

Finding a Needle in a Haystack: The Effects of Searching and Learning on Pick-Worker Performance

Robert J. Batt

Wisconsin School of Business, University of Wisconsin-Madison, Madison, WI 53706, bob.batt@wisc.edu

Santiago Gallino

Tuck School of Business, Dartmouth College, Hanover, NH 03755, santiago.gallino@tuck.dartmouth.edu

The rise in online and multichannel retailing has pushed retailers to give increased attention to their order fulfillment operations. We study “chaotic storage” fulfillment systems in which dissimilar items are stored together in a single location. This necessitates a *searching* task as part of the picking process which has not been previously studied. We show that pick times increase by as much as 16% as the searching task becomes more difficult. However, the deleterious effect of searching decreases with pick worker experience. Using simulation, we show that pick times can be improved by incorporating distance, bin density, and picker experience into pick assignments and pick routing. Through properly combining the details of the task and the workers, order fulfillment productivity can be increased by approximately 5%.

Key words: Retail operations, Warehouse operations, Learning curve, Task assignment, Empirical studies.

History: Accepted for Management Science, Jan. 2, 2018

1. Introduction

Online retail sales in the United States represented 3.2% of total retail sales during the first quarter of 2007. By 2016, online sales had increased to 7.8% of total retail sales and was still growing (U.S. Census Bureau 2016). In many retail segments, such as books, electronics, and toys, online sales are already above 20% of total sales (Kapner 2015). One result of this growth is an essential change in the retailer’s back-end operations. While retail warehouses have historically been focused on shipping cases to stores, they are now becoming order fulfillment centers shipping single items to individual customers. The retailer’s ability to efficiently fulfill thousands of individual orders is an essential part of success in a multichannel retail world. This new challenge provides motivation for the present paper. In particular, we study the retail order picking process and we show how retailers can improve their operations by better understanding and managing the drivers of this process.

Warehouse operations has been a popular area of Operations Management research for decades (see de Koster et al. (2007) and Gu et al. (2010) for surveys of the literature). Much of this work has focused on management of pallets, cases, and unit loads, as companies distribute goods down

the supply chain (Manzini 2012). However, driven by the rapid rise of e-commerce over the past 15 years, companies have been putting increased focus on picking individual items for shipment directly to end-users. Online customers value fast deliveries at low cost (Fisher et al. 2016). Retailers that want to remain competitive need to react to this demand.

As noted by Bartholdi and Hackman (2014), “Piece-picking is the most labor intensive activity in the warehouse because the product is handled at the smallest units of measure. Furthermore, the importance of piece-picking has greatly increased because of pressure to reduce inventory while expanding product lines. Warehouses that, 20 years ago, might have shipped cartons to customers now ship pieces and much more frequently.” Researchers have estimated that order-picking accounts for 55% to 65% of total warehouse operating expenses. (Coyle et al. 1996, de Koster et al. 2007).

We study the implications for the retailer’s operations of using “article mixed placing” (Gudehus and Kotzab 2012), or more colloquially referred to as “chaotic storage” (Reichardt 2011), in which inventory is not stored together with like items in predefined locations, but rather is split up and assigned to any space that is available and with different stock keeping units (SKUs) being stored together in the same location or bin. For example, a storage bin at a Zappos.com order fulfillment center (OFC) might hold 15 to 20 different pairs of shoes, each a different SKU (Frei et al. 2011). Chaotic storage is used by retailers selling widely varying product types such as Amazon.com, Walmart, Macy’s, and New Balance (Reichardt 2011, Author 2016), as well as by third-party logistics companies such as Quiet Logistics (Author 2016), and by single-product-category companies, such as our research partner company selling women’s apparel.

The key benefit of chaotic storage is the efficient use of space, leading to reduced total space needs. This is a relevant factor because in an effort to reduce delivery time, OFCs are often located on the outskirts of major cities near demand where real estate (and thus warehouse rental) is relatively expensive. At the same time, most online retailers experience high return rates. Return rates between 20 and 30 percent are common for online apparel retailers (Banjo 2013, Chao 2015, Kapner 2016). In a dedicated storage location system, a great deal of empty space needs to be maintained to allow for returned inventory to be restocked. In the chaotic system, this is not an issue because the empty space can be used by another item right away, and a returned item can go in whatever space is available at the time of its return.

However, chaotic storage adds a layer of complexity to the picking process because it requires the picker to search for the correct item within the bin once she arrives at the right location. The impact of this intra-bin searching on pick time (the total time to walk to the item, find it in the bin, and physically remove it) and productivity has not been previously studied.

Based on data from an e-commerce apparel retailer, we perform a detailed econometric study of the impact of intra-bin searching on pick times. In addition, using simulation we show how the

empirical results can be used to improve pick assignment, routing, and productivity. Specifically, we make the following four contributions:

1. We show that pick time increases with both the number of items and the density of items in the bin (i.e., items per foot). For example, we find that pick time increases by 16% as density varies over its typical range of 9 items per foot to 22 items per foot. We also confirm that pick time increases with the required travel distance, with an additional foot of travel distance leading to a 0.21 second increase in pick time.

2. We show that pick times decrease as workers gain picking experience. We find that mean pick time drops by 4.2% with each doubling of a picker’s experience level. Further, we show that this pick time improvement comes via an improvement in the intra-bin searching portion of the task and not from an improvement in the travel portion of the task. We show that the detrimental marginal effect of bin density drops by 86% as a picker progresses from novice to experienced, while the marginal effect of distance remains unchanged with experience.

3. We show that workers exhibit heterogenous learning, and that the variance in performance across pickers diminishes with experience. However, due to the heterogenous learning, a picker’s performance as a novice is not necessarily indicative of his relative performance at higher levels of experience.

4. We show that incorporating bin density into pick routing heuristics can improve pick times by 0.5% to 7%, depending on worker experience. Further, we show that incorporating pickers’ experience into pick assignments (i.e., strategically assigning orders across pickers with heterogeneous experience) can improve productivity by an additional 0.5% to 3%.

These findings together show that intra-bin searching has a meaningful impact on pick worker productivity and that ignoring this effect leads to sub-optimal pick assignment and routing. Our work shows that managers can temper the deleterious effects of intra-bin searching through incorporating density and experience into pick routing and task assignment.

2. Literature Review

In the introduction to an exhaustive review of the literature on warehouse operations, Yavuz Bozer notes that the growth of online retailing has “fundamentally changed” the way companies see their distribution centers not only from the operational perspective but also thinking of them as profit-centers rather than cost-centers (Manzini 2012). Picking, packing and shipping the right items to individual customers is a completely different endeavor than the traditional fulfillment of bulk orders to serve brick and mortar stores. The level of accuracy and the speed at which the order is completed at the fulfillment center is now extremely relevant.

Gu et al. (2007) provide an organizing framework of warehouse design and operation. As part of this framework, they show that one of the key decisions in defining the warehouse operating strategy

is selecting where items will be stored, which is known as the “storage location assignment problem” (SLAP). Generally, the SLAP seeks to reduce material handling cost and space requirements. Many location assignment policies have been studied, and [Turner \(2009\)](#) provides a summary of the most common policies.

The two most fundamental storage allocation policies are dedicated assignment and random assignment, while most other policies are variations of these two ([Hausman et al. 1976](#)). Under dedicated assignment, each stock keeping unit (SKU) is assigned a fixed storage location for the duration of the planning period (see [Goetschalckx and Ratliff \(1990\)](#) for an interesting work on the duration-of-stay policy). In contrast, under random assignment, incoming items are assigned randomly to any available location. This policy allows the same product to be allocated to multiple different locations in the OFC. The main advantage of dedicated assignment is that workers become familiar with where items are located. A disadvantage is that dedicated assignment has very high space requirements as sufficient space must be allocated for the maximum inventory level of each SKU ([de Koster et al. 2007](#)). Conversely, random assignment provides very high space utilization since empty space can be immediately reassigned to incoming product. However, random assignment requires more sophisticated information systems to track the location of all items. Neither assignment system has an inherent advantage with regard to pick worker travel distance ([de Koster et al. 2007](#)).

Other assignment policies, such as class based, turnover based, and duration of stay based policies seek to improve on the dedicated and random policies by using aspects of each policy and by incorporating information about product demand to assign locations so as to minimize travel distance. We direct the reader to [Gu et al. \(2007\)](#) and [Turner \(2009\)](#) for further information on storage allocation policies and related references.

Chaotic storage is a modified form of random storage in that incoming items are assigned randomly to available space, but in chaotic storage different SKUs can be stored within the same location or bin. This provides even greater flexibility and higher space utilization than regular random storage assignment. However it adds complexity in that the pick worker must not only find the correct location or bin but then must search within the bin for the correct item. This intra-bin searching is the focus of our work.

The warehouse design and operation framework of [Gu et al. \(2007\)](#) also highlights the order-picking strategy as a key design decision. This involves decisions regarding whether the picker will travel to the items (picker-to-parts) or the parts will travel to the picker (via an automated system, parts-to-picker), and how picks will be assigned to pick workers (e.g., pick all items in a single order, pick all items in a single physical zone). The literature also focuses on algorithms and heuristics to minimize the distance traveled by the picker to complete an order or a set of orders. [Petersen and Aase \(2004\)](#) examine the effects of picking, storage, and routing on order picker travel

and emphasize the cost implications of these activities for order fulfillment. As with the SLAP, many systems are used in practice and have been studied in the literature. We refer the reader to [de Koster et al. \(2007\)](#) and [Manzini \(2012, Chapter 6\)](#) for an overview of this literature.

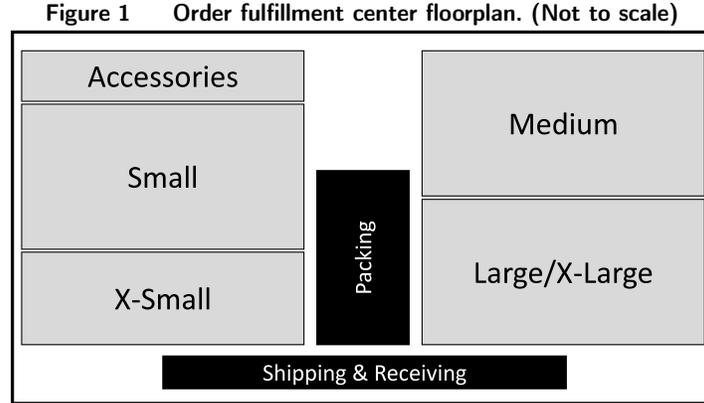
Prior literature on picking strategy generally assumes that the search time the picker needs to find an item once she arrives at the item's location is negligible. In other words, prior work has accounted for travel time but not for intra-location searching time. Ignoring this search time is reasonable when every location contains only a single SKU, but as we show in this paper, this simplification is problematic when multiple SKUs are in a single location.

One exception to the negligible search time assumption is the literature on “miniload” systems. A miniload system is an automated, chaotic storage, parts-to-picker system in which bins or totes containing one or more SKUs are robotically delivered to pickers for picking of the correct items ([Johnson and Brandeau 1996](#)). The literature on these systems is analytical in nature and focuses on optimal system design, generally assuming deterministic, exponential, or uniform distributed search times that are independent and identically distributed for all picks (e.g. [Bozer and White 1990](#), [Foley and Frazelle 1991](#), [Gibson and Sharp 1991](#)). To the best of our knowledge, ours is the first paper to empirically study the search time of a chaotic storage system.

Because the phenomenon of interest in our work involves workers searching and learning, our work also connects with the behavioral operations management literature. As [Boudreau et al. \(2003\)](#) note, behavioral OM research addresses “how human considerations affect classical OM results.” The literature on behavioral operations includes a wide range of operations management topics ranging from supply chains ([Beer et al. 2014](#), [Davis et al. 2014](#)), to demand forecasting ([Kremer et al. 2011](#), [Tong and Feiler 2016](#)), to contracting ([Becker-Peth et al. 2013](#)), to queuing ([Kremer and Debo 2015](#), [Batt and Terwiesch 2015](#)). [Bendoly et al. \(2006\)](#) provides a detailed literature review.

Within this stream of literature, the topic of learning from experience has had a prevalent role. [Dutton and Thomas \(1984\)](#) provide an overview of early work on learning in production settings. More recent work has focused on issues such as team learning ([Huckman et al. 2009](#), [Tucker et al. 2007](#)), learning from others ([KC et al. 2013](#), [Song et al. 2015](#)), drivers of learning effectiveness ([Mukherjee et al. 1998](#), [Lapr e et al. 2000](#)), and the impact of task variety on learning ([KC and Staats 2012](#), [Ramdas and Randall 2008](#)).

Our work is also relevant beyond the OFC and learning literature. Retailers are becoming “omnichannel” retailers ([Bell et al. 2014](#)), hence, order-fulfillment functions are extremely relevant and are now also taking place inside the brick and mortar (B&M) stores. For example, many retailers now offer “buy online, pickup in store” or “ship from store” options, and these require a store employee to act as a picker and pick an item from the sales floor (e.g., [Gallino and Moreno 2014](#), [Gao](#)



and Su 2016). Our work highlights the importance of accounting for the time-cost of searching for an item, which can help retailers staff appropriately to meet both B&M and online demand.

Our work builds on and contributes to the above mentioned streams of literature. We show that learning occurs in the item picking process, and further, we decompose the learning effect into its component effects on the travel and searching sub-tasks. We contribute to the order picking literature by showing that picker productivity can be improved by incorporating the effects of searching and learning into pick assignment algorithms.

3. Empirical Setting & Data Description

Our data come from an online women’s apparel retailer. We focus on its main OFC in California. At any given time, this 100,000 square foot facility has a total inventory of approximately 180,000 items (i.e., articles of clothing) and ships an average of 20,000 items per day. The data we collected from the company covers the period May 2014 through January 2015.¹

The raw data consists of timestamps of every item event (e.g., movement or status change), the key events being when incoming inventory is placed in a bin, and when an item is removed from a bin to be added to a shipment. The item event data also include an employee identifier code for the employee responsible for the event, and a bin location code for events involving the placing or picking of items from the OFC floor. From this raw data we generate the analysis dataset which consists of an observation for every item that is picked from the floor for shipment to a customer.

The OFC is organized into five inventory-holding zones by item type and size (Figure 1). The Accessories zone primarily contains small, non-apparel items, such as jewelry and handbags, stored in drawers or in small bags on hanging rods. The other four zones contain only apparel on hangers, such as pants and blouses, and are organized by size (extra small, small, medium, large/extra-large).

¹ We drop two weeks of data in early September because an error in the company’s inventory tracking system led to erroneous data. While we do not know exactly how many picks are lost, this represents about 5% of the observation period, which would suggest about 220,000 picks are lost.

Since all items in these zones are on hangers, the zones consist entirely of rows of hanging racks. These racks are divided by hanging signs into segments or “bins” of between 1.4 and 3.3 feet long containing an average of 47 items (standard deviation of 15). Each size zone is laid out as a grid of 1,000 bins, each with a unique identifier. The remainder of the OFC floor contains space for folding and packing the outgoing items, for processing incoming inventory (new and returns), and for supervisor workspace.

The company’s business model dictates that virtually all shipments consist of exactly five items. Further, 99% of shipments contain at most one item from the accessory zone, with the rest coming from the hanging-items zones (Accessories are picked via a different process that does not record timestamps at the time of the pick. Therefore, for the remainder of the paper, we focus only on apparel goods picked from the four size-based zones.). We take advantage of this structure to identify the factors that affect pick time. The key benefit of this system for our analysis is that all of the apparel (non-accessory) items in a shipment are picked sequentially by a single worker without any other intermediate tasks (i.e., a worker picks each item in sequence for a single shipment before moving on to the items for the next shipment). Thus, the data allow us to directly observe the time and location of each pick of each shipment, from which we calculate the minimum necessary travel distance and the time between picks².

The raw item-picking events are created as the pick worker, or “picker” moves about the OFC picking items for shipments. The picker pushes a laptop computer on a wheeled cart that provides the item number, a photograph, and the bin number of the next item to be picked. The picker walks to the prescribed bin by a route of her own choosing. Since bins contain many different items, after arriving at the bin the picker must search through the items in the bin to find the correct item. By design, there are no duplicate SKUs in a single bin, thus, the picker must find the one item in the bin that meets the need. Once the correct item is found, the picker picks the item off the rack, scans the attached barcode, and places the item in a box on the cart. Scanning the item generates the “picked” item event in the database and triggers the picker’s computer to display the information for the next item to be picked. The picker continues on in this way, item after item, periodically returning to the central Fold and Pack area to drop off picked items and get new pick lists to fulfill.

While the actual travel path taken by the picker is not observed, by knowing the starting and ending location of a pick and the physical layout of the rows and cross aisles of the OFC, we can calculate the minimum necessary travel distance, which is actually preferred for our purposes. We

² At the time of the data collection, the company was not performing any routing optimization. Pick lists were sorted numerically by the bin numbers of the required items (e.g., Bin 50, 105, 235, and 788), and pickers could choose to work in either ascending or descending order through the pick list. We did not observe any deviation from this picking approach in our time on the floor.

use a modified form of “taxicab geometry,” sometimes referred to as “urban geometry,” to calculate the travel distance (Krause 2012, Malkevitch n.d.). This form of geometry assumes that travel can only occur on a grid, which in our case is formed by the aisles and cross-aisles between the racks of clothes. We also take into account that pickers must walk around the Packing zone when they cross from one side of the OFC to the other.

We observe the number of items present in a bin at the moment an item is picked from the bin, which we refer to as the “bin load,” or simply “load.” We also know the nominal physical size (i.e., length in feet) of each bin, and thus calculate the density of items in the bin as the bin load divided by the bin size. Table 1 provides summary information about the item and pick variables described above.

Table 1 Descriptive statistics for picking a single item

Variable	Mean	Standard Deviation
Time (sec.)	49.2 (0.02)	30.7
Distance (ft.)	86.5 (0.04)	65.4
Bin load (items)	46.9 (0.01)	15.2
Bin density (items/ft.)	15.7 (0.00)	4.5
N	2,427,758	

Standard errors in parentheses

We note that while bin density is one measure of search complexity, another possible driver of search complexity is the visual salience of the desired item relative to the other items in the bin. For example, a green dress in a bin full of black pants is highly salient and easy to find, while picking a specific pair of black jeans from that same bin is much more difficult due to the lack of visual salience. We examine this issue in Appendix A and validate that bin density is a good predictor of search complexity.

Because each item event record contains an employee identifier, we can calculate a worker experience variable: the cumulative number of items picked by each worker at the moment she performs each pick. This variable allows us to estimate the effect of cumulative picking experience. Table 2 provides descriptive statistics of this Pick Count measure as well as of overall picker experience and tenure over the nine months of the dataset. We note that while this is certainly a high turnover job, we do observe many pickers that stay in the job long enough to pick tens of thousands of items.

Table 2 Picker Experience Descriptive Statistics

Variable	Description	Mean	Standard Deviation	Percentile			N
				5th	50th	95th	
Pick Count	Picker experience at time of pick	11,549	12,526	565	7,476	36,474	2,427,758
Picker Experience	Total number of picks by picker	10,627	11,825	1,256	6,872	32,955	526
Picker Tenure	Total days worked by picker	45.8	37.1	12	34	120	526

Descriptive statistics for 526 pickers.

4. Model Formulation

Our empirical analysis focuses on estimating the impact of factors such as distance, load, density, and experience on pick time. Since this analysis is all executed as variations on a single econometric model we begin by describing the base model and then proceed to present the analysis of each variable of interest in turn in Section 5. The dependent variable of interest is the pick time ($PICK_TIME_{ij}$) which is defined as the elapsed time in seconds between the completion of successive picks (pick $i - 1$ and pick i) by picker j . Because our dependent variable is a duration, we choose a survival model for our analysis. Specifically, we use a parametric accelerated-failure-time (AFT) model which relates the log of a duration to a vector of covariates and a random error term ϵ through a linear equation (Greene 2012, Sec. 19.4.3). The base model takes the form

$$\ln(PICK_TIME_{ij}) = \alpha_j + \eta_1 DIST_{ij} + \eta_2 LOAD_{ij} + \eta_3 SIZE_INV_{ij} + \eta_4 DENSITY_{ij} + \mathbf{Z}_{ij}\boldsymbol{\theta} + \epsilon_{ij} \quad (1)$$

$DIST_{ij}$ is the minimum necessary distance for picker j to travel from pick $i - 1$ to pick i and, as described in Section 3, is computed using urban geometry. $LOAD_{ij}$ is the number of items in the target bin at the time of pick i . $SIZE_INV_{ij}$ is the multiplicative inverse (reciprocal) of the size of the bin. $DENSITY_{ij}$ is the number of items per foot and is calculated as

$$DENSITY = \frac{LOAD}{SIZE} = LOAD \times SIZE_INV. \quad (2)$$

Parameterizing the load and size variables in this way allows density to be included in the model as a multiplicative interaction term of $LOAD$ and $SIZE_INV$. This allows for a natural interpretation of the marginal effect of $LOAD$ because this formulation of the model recognizes that adding an additional item to the bin not only increases the load, but also changes the density of the bin, holding the size of the bin constant.

\mathbf{Z}_{ij} is a vector of time related control variables including month, day of the week, and hour of the day. α_j is a picker specific fixed effect.

We test several common distributional assumptions for ϵ_{ij} and find that a logistic distribution provides the best fit based on the Bayesian Information Criterion. Note that this model is commonly

referred to as a loglogistic AFT model because AFT models are generally named for the distribution of τ where $\tau = e^\epsilon$, and thus τ is distributed as loglogistic (Cleves et al. 2010). ϵ_{ij} is assumed to have mean zero and standard deviation of $\frac{\pi\gamma}{3}$. γ is estimated as an auxiliary parameter of the survival model. Other distributional assumptions for ϵ_{ij} give qualitatively similar results.

Because we are interested in understanding the drivers of pick time for typical pickers under typical conditions, we make a few restrictions to the data used in our analysis. The full data contains about 4 million picks. We restrict the sample to shipments that contain exactly five items, with four or five of those items being apparel, and all of the items being picked by a single worker (drop 481,819 picks). These types of orders comprise 88% of the data. We further restrict the sample by dropping picks performed by employees who work a total of 10 or fewer days in the OFC so as to not bias the results with these pickers who either did not stay with the company very long or who did not have picking as their primary role (drop 285,820 picks). In robustness checks, we include these low-experience workers and find no meaningful differences in the results. We also drop observations with a pick time greater than 600 seconds (95th percentile is 107 seconds) because these observations likely represent anomalous situations such as a missing item, or an interruption, which are outside our interests (drop 2,922 picks). Lastly, we drop picks that are not immediately preceded by a pick from the same shipment (drop 736,730 picks). We do this because to calculate travel distance we need to know the starting and stopping location of the picker, and we only observe the location of the picker at the moment they pick an apparel item off the rack. The final analysis dataset includes 2,427,758 pick events by 526 unique workers.

5. Analyzing Drivers of Pick Time

We start our analysis by examining the effects of the physical characteristics of the pick on the pick time. We then progress to examine how learning through experience impacts pick times. Lastly, we examine the interaction of the physical characteristics and pickers' experience to better understand which parts of the picking task are most impacted by experiential learning.

5.1. Physical Drivers: Distance, Load, and Size

We first examine how the physical characteristics of the required pick impact pick time. Pick time is measured from the time of one pick to the next and consists of two tasks: walking and searching. Certainly, we expect distance to have a positive impact on pick time as walking farther generally requires more time ($\eta_1 > 0$).

Because of the interrelated nature of bin load, size, and density, it is more difficult to articulate the marginal effects. In practice, the bin size remains essentially fixed while the bin load and resulting density vary as items move on and off the rack. We expect pick time to increase with load because the picker is looking for precisely one item in the bin and finding this single item likely takes longer

when there are more items in the bin. Furthermore, because identifying the correct item involves visually matching the item to the provided item number and picture, we expect this matching to be impeded by higher density racks as it is more difficult to see the items and access the item barcode tags. Thus, we expect both η_2 and η_4 of Equation 1 to be positive and the marginal effect of *LOAD* to measure the combined effect. The marginal effect of bin size (and thus *SIZE_INV*) is less clear. Modest changes in bin size likely have little effect on pick time apart from the resulting change in density. Thus we anticipate that η_3 will be close to zero. The marginal effect of *SIZE_INV* should be positive as it incorporates the impact of size on density (an increase in inverse size represents a decrease in size and an increase in density).

Table 3 Effects of Physical Factors

	1	2	3
DISTANCE		0.0045*** (0.0001)	0.0045*** (0.0001)
LOAD		0.0014*** (0.0004)	
SIZE_INV		0.0006 (0.0330)	
DENSITY		0.0065*** (0.0011)	0.0109*** (0.0004)
Picker FE	Yes	Yes	Yes
Time Controls	Yes	Yes	Yes
Gamma	0.2823 (0.0021)	0.2200 (0.0023)	0.2201 (0.0023)
N	2,427,758	2,427,758	2,427,758
BIC	3,600,002	2,496,628	2,497,568

Cluster robust standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Table 3 provides the results of estimating Equation 1 without and with the physical driver factor variables (Models 1 and 2, respectively). The results show that distance has a positive impact on pick time. When exponentiated, these coefficients can be interpreted as time ratios for a 1-unit change in the related variable. Thus a 1 foot increase in distance is associated with a 0.45% increase in pick time ($\exp(0.0045) = 1.0045$). The mean marginal effect of 1 foot of distance is 0.21 seconds. We note that this corresponds to a walking speed of 3.3 miles per hour (4.8 feet per second) which is close to the normal adult walking pace of 3.1 miles per hour (Rose and Gamble 2006).

As expected, bin load and density both have a positive impact on pick time. The mean marginal effect of load, which incorporates both the load and density effects, is 0.16 seconds per item. The main effect of inverse bin size is not statistically significant, as expected, but the marginal effect, which incorporates the density effect, is positive and significant, as expected.

5.2. Physical Drivers: Distance and Density

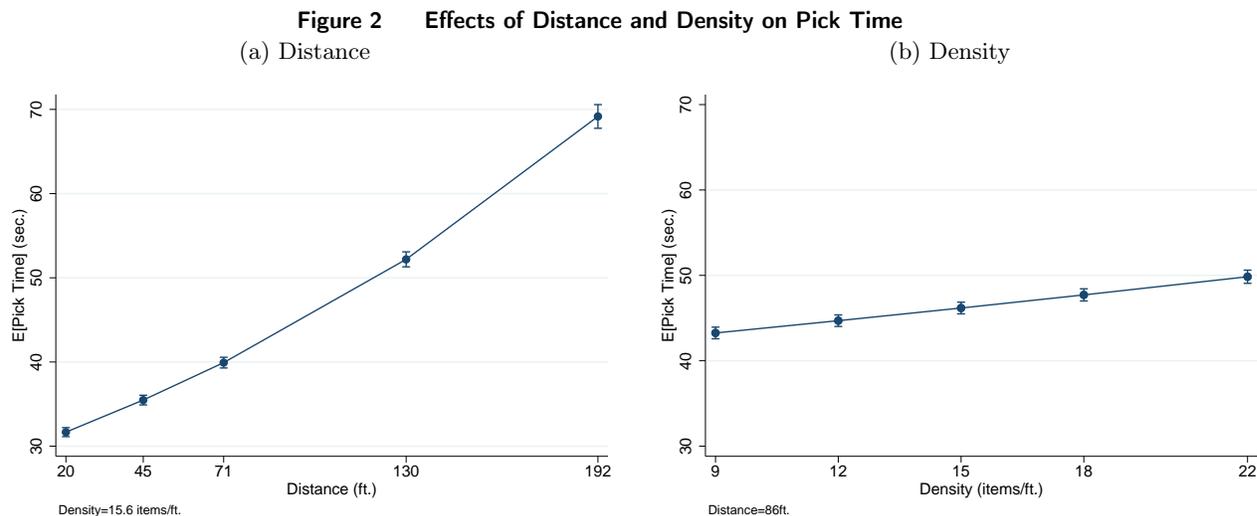
Bin load, size, and density collectively impact the difficulty of picking the correct item. However, an alternative and more direct way to model this effect is to include only bin density, because the bin density captures both the number of items in the bin and the size of the bin. We modify Equation 1 as follows:

$$\ln(\text{PICK_TIME}_{ij}) = \alpha_j + \beta_1 \text{DIST}_{ij} + \beta_2 \text{DENSITY}_{ij} + \mathbf{Z}_{ij}\boldsymbol{\theta} + \epsilon_{ij} \quad (3)$$

This formulation has the advantages of being more parsimonious and the marginal effect of bin density can be estimated directly, rather than as an aggregate effect with bin load or size.

The results of estimating Equation 3 are shown in Model 3 of Table 3. The coefficient on *DISTANCE* is unchanged from Model 2, and the coefficient on *DENSITY* is positive and significant. Exponentiating the *DENSITY* coefficient shows that the mean marginal effect of increasing the bin density by 1 item per foot is a 1.1% increase in pick time, or approximately 0.51 seconds. The Bayesian Information Criterion (BIC) indicates that Models 2 and 3 provide similar levels of fit to the data. For the remainder of the analysis we build on the more parsimonious Model 3, however all results hold if Model 2 is used.

Figure 2 shows predicted pick times over typical ranges of distance and density and helps put the marginal effects of these variables into context. Figure 2a shows that as distance ranges from 20 feet to 192 feet (5th to 95th percentile), expected pick time ranges from 35 seconds to almost 75 seconds. Figure 2b shows that as density varies over its typical range, expected pick time ranges from about 43 seconds to 50 seconds, an increase of 16%.



Note: Error bars indicate 95% confidence interval

In summary, picking time is comprised of walking and searching, and thus is affected by the required travel distance, the number of items through which the picker must search, and the density of those items in the bin. While typical pick planning focuses on reducing distance, our results suggest that pick planning should also take into consideration bin density because there may be times when a faster pick time can be achieved by routing to a less full bin even if it is not the closest bin. In Section 6 we examine the benefits of incorporating bin density into pick routing decisions.

5.3. The Role of Experience

5.3.1. Average Effects of Experience We now turn to the topic of experience and learning. Specifically, we examine if pick times improve with picker experience. We do this by estimating the following survival model, which is a modification of Equation 3.

$$\ln(PICK_TIME_{ij}) = \alpha_j + \beta_1 DIST_{ij} + \beta_2 DENSITY_{ij} + \beta_3 \ln(PICK_COUNT_{ij}) + \mathbf{Z}_{ij}\boldsymbol{\theta} + \epsilon_{ij} \quad (4)$$

$PICK_COUNT_{ij}$ is the number of picks picker j has made prior to pick i . Following common practice from the literature on learning (e.g., [Pisano et al. 2001](#), [Staats and Gino 2012](#)), we take the natural log of $PICK_COUNT$. Our parametric survival model has a logged dependent variable, which is also standard in the learning literature. This has two benefits. First, it allows for a functional form in which the benefits of experience are greatest early on and then taper off. Second, it eliminates the problem of censored data (pickers only being observed in the time window of the dataset, but not before) by controlling for experience as a proportional effect rather than an effect in levels ([Lapr e and Tsiriktsis 2006](#)).

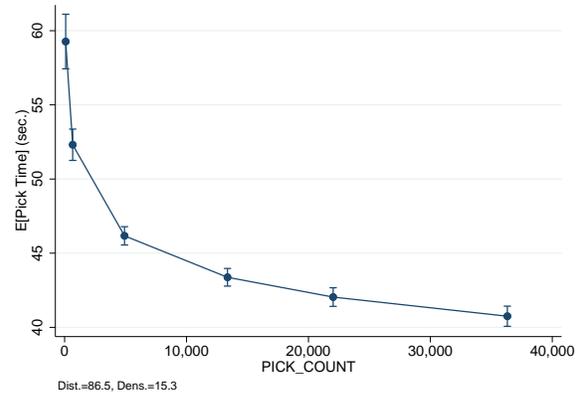
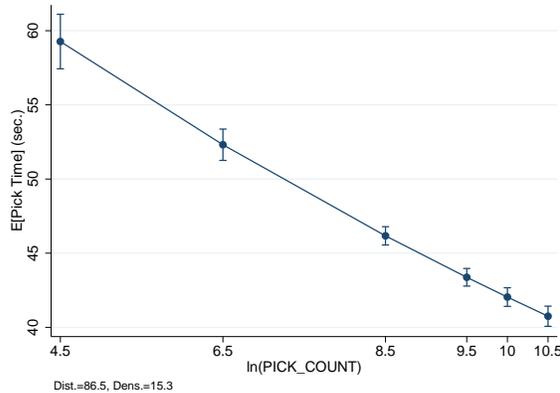
The estimation results of Equation 4 are provided in Model 1 of Table 4. The negative coefficient on $\ln(PICK_COUNT)$ indicates that pick times improve with picking experience. A 1 unit increase in $\ln(PICK_COUNT)$ leads to a 6.0% reduction in pick time ($\exp(-0.0624) = 0.94$). Alternatively, a doubling of $PICK_COUNT$ leads to a 4.2% reduction in pick time ($2^{-0.0624} = 0.958$). Figure 3 shows predicted pick times as a function of both $\ln(PICK_COUNT)$ and $PICK_COUNT$. Recall from Table 2 that 50% of picks are performed by pickers with prior experience of at least 7,476 picks ($\ln(PICK_COUNT) = 8.9$), and 5% of picks are performed by workers with experience of at least 36,474 prior picks ($\ln(PICK_COUNT) = 10.5$).

The above result shows that the effect of increased experience is a reduction in pick times. Now we explore how the learning affects the two sub-tasks involved in a pick: walking to the bin and searching for the item in the bin. We want to understand whether the net learning effect observed in Figure 3 and Model 1 of Table 4 comes from improvements in the walking task (e.g., navigating

Table 4 The Effect of Experience on Pick Time

	(1)	(2)	(3)
DISTANCE	0.0045*** (0.0001)	0.0022*** (0.0002)	0.0029*** (0.0002)
DENSITY	0.0110*** (0.0004)	0.0111*** (0.0025)	0.0454*** (0.0023)
ln(PICK_COUNT)	-0.0624*** (0.0033)	-0.0877*** (0.0059)	
ln(PICK_COUNT)xDIST		0.0003*** (0.0000)	0.0002*** (0.0000)
ln(PICK_COUNT)xDENS		-0.0000 (0.0003)	-0.0039*** (0.0003)
Picker FE	Yes	Yes	Yes
Time Controls	Yes	Yes	Yes
Gamma	0.2188 (0.0023)	0.2216 (0.0040)	0.2188 (0.0023)
N	2,427,758	2,427,758	2,427,758
BIC	2,470,016	2,560,981	2,470,454

Cluster robust standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$ **Figure 3** Effects of Experience(a) $\ln(PICK_COUNT)$ (b) $PICK_COUNT$ 

Note: Error bars indicate 95% confidence interval

around the warehouse), the searching task, or both. If experience leads to pickers getting better at finding their way from one bin to the next, then we would expect the marginal effect of an additional foot of distance to be lower for more experienced pickers. Likewise, if pickers get better at identifying and selecting the correct item in the bin, we would expect the marginal effect of an additional item in the bin to decrease with experience.

To examine these effects we add two interaction terms to Equation 4 as follows:

$$\begin{aligned}
\ln(PICK_TIME_{ij}) = & \alpha_j + \beta_1 DIST_{ij} + \beta_2 DENSITY_{ij} + \beta_3 \ln(PICK_COUNT_{ij}) \\
& + \beta_4 \ln(PICK_COUNT_{ij}) \times DIST_{ij} \\
& + \beta_5 \ln(PICK_COUNT_{ij}) \times DENSITY_{ij} + \mathbf{Z}_{ij} \boldsymbol{\theta} + \epsilon_{ij}
\end{aligned} \tag{5}$$

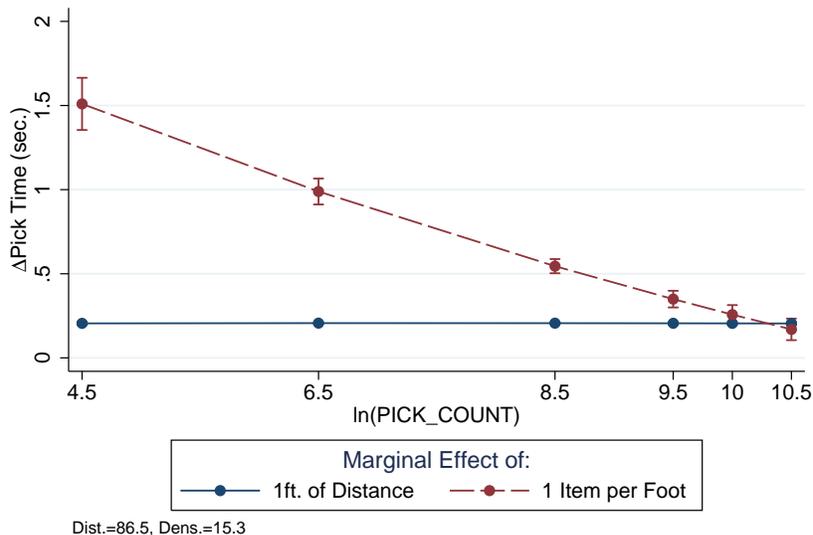
The $\ln(PICK_COUNT_{ij}) \times DIST_{ij}$ term allows for learning in the walking portion of the picking task, and the $\ln(PICK_COUNT_{ij}) \times DENSITY_{ij}$ term allows for learning in the searching portion of the picking task.

Equation 5 includes both the main effect of pick count (β_3), as well as interactions with distance and density (β_4, β_5). This formulation allows experience to have an effect on pick time both by itself and via interactions. However, this is not necessarily an accurate reflection of the actual picking process. Since the picking process is dominated by walking and searching, we might want to restrict experience to only having an effect via distance and density. This is done by omitting $\ln(PICK_COUNT)$ from Equation 5. This has the further implication that the magnitude of the benefit of experience diminishes when distance and density are close to zero. We believe this model is a better representation of the effect of experience on performance, however, for completeness we present both models.

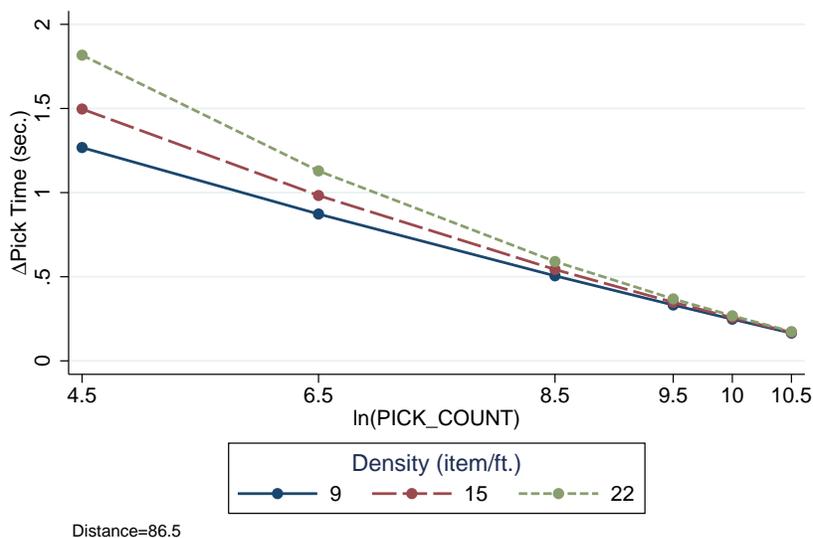
The results of estimating Equation 5 with and without the $\beta_3 \ln(PICK_COUNT)$ term are displayed in Models 2 and 3 of Table 4, respectively. In Model 2, we see that the interaction of experience and distance is positive and significant while the interaction of experience and load is nonsignificant. In Model 3, we see that all the coefficients are unchanged in sign and significance, with the exception of the $\ln(PICK_COUNT) \times DENSITY$ coefficient which becomes negative and strongly significant. The BIC indicates that this model provides a markedly better fit than does Model 2, so we focus on Model 3. Since interpreting marginal effects of interaction coefficients in nonlinear models is difficult by inspection, as recommended by Greene (2012, Sec. 7.2.5), we turn to predicted marginal effects plots to better convey the results.

Figure 4 shows the marginal effects of distance and density as functions of experience. The solid line indicates the marginal effect of an additional foot of distance on pick time. Given that the coefficient of the $\ln(PICK_COUNT) \times DIST$ interaction is positive (Models 2 and 3 of Table 4), one might expect this line to be upward sloping. However, due to the small magnitude of the coefficient (0.0002), we see that the marginal effect is essentially constant at 0.2 seconds per foot over the observed range of experience. The point estimates vary by less than 0.002 seconds per foot and are statistically indistinguishable. Thus for all practical purposes, the marginal effect of distance is unchanging with experience.

The dashed line in Figure 4 indicates the marginal effect of an additional item per foot of bin on pick time. In contrast to distance, this effect drops from 1.5 seconds per item per foot to 0.2 seconds

Figure 4 Marginal effects of distance and density with experience

Note: Error bars indicate 95% confidence interval

Figure 5 Marginal effect of density is largest for heavily loaded bins

per item per foot as picker experience ranges from low to high, an 86% reduction. To convert this marginal effect of density (items per foot) to the marginal effect of bin load (items per bin) we divide by the median bin size of 3.1 feet to see that the marginal effect of an additional item in a bin ranges from 0.48 seconds per item to 0.05 seconds per item. Thus, we find that when we assume that the effects of learning occur via distance and density, the benefits of learning come entirely from improvements in searching for the right item and not from improvement in navigating to the correct bin.

Figure 5 decomposes the learning and density effect even further by plotting the marginal effect of an additional item per foot in the bin for three levels of density. We see that not only do novices exhibit a relatively large marginal response to density, but that this effect is exacerbated for bins that are heavily loaded (the top line of the graph). As pickers gain experience, they become less impacted by density. The improvement in performance converges such that for an experienced picker, the marginal effect of density is essentially the same regardless of the bin density.

In summary, we show that picking experience leads to improved pick time performance and that this performance gain comes from an improvement in the search portion of the picking task, while there is little change in the walking portion of the task. These results suggest that if a company has a workforce that is heterogeneous in picking experience, it might be beneficial to take worker experience into account when assigning picks and pick routes to workers. We examine this in Section 6.

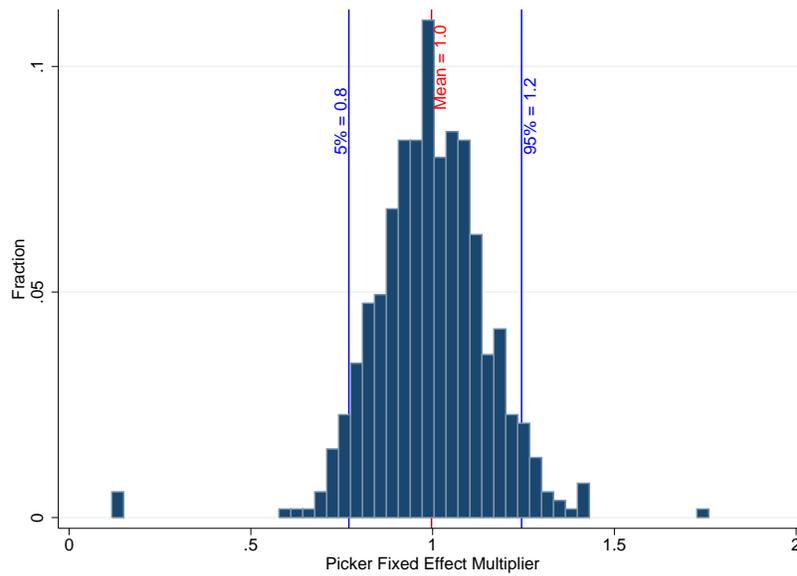
5.3.2. Individual Effects of Experience As described in Section 4, we include a picker fixed effect (α_j) in all the models we estimate. This controls for performance differences between workers that are constant over time. Thus, all the results described so far are “on average” results, controlling for the picker fixed effects. However, it is also interesting to examine the differences between workers as captured by the fixed effects, and more specifically, whether the effects of experience are the same across pickers.

To get a sense of the range of performance across pickers we look at the distribution of the fixed effects estimated by Equation 4 (Model 1 of Table 4). To make interpretation easier, we exponentiate each estimated α_j to yield pick time multipliers. For example, a picker with $\alpha = 0.1$ has a pick time multiplier of $\exp(0.1) = 1.105$ and therefore has pick times about 10.5% longer than the omitted reference picker.

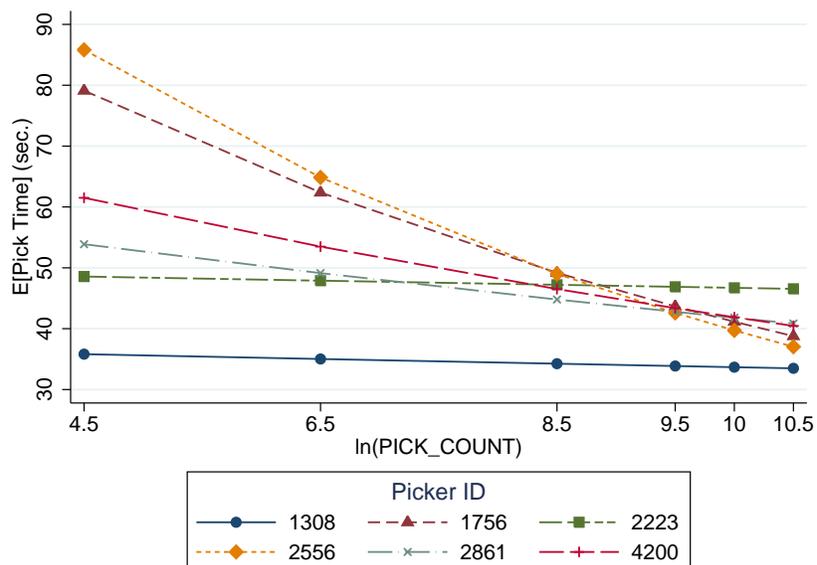
Figure 6 displays the histogram of the picker pick time multipliers relative to a “middle of the road” omitted reference picker. We see that while there is a great deal of variation among picker performance, 90% of the 526 pickers have pick time multipliers within 20% of the reference picker.

Equation 4 assumes that learning is homogeneous across pickers and is captured by the $\beta_3 \ln(PICK_COUNT_{ij})$ term of the model. However, it is possible that learning is in fact heterogeneous, and each picker learns at a different rate (Gans et al. 2010). To estimate this, we modify Equation 4 to also include interaction terms between an indicator function for picker j ($\mathbb{1}_j$), and the experience variable $\ln(PICK_COUNT_{ij})$.

$$\begin{aligned} \ln(PICK_TIME_{ij}) = & \alpha_j + \beta_1 DIST_{ij} + \beta_2 DENSITY_{ij} + \beta_3 \ln(PICK_COUNT_{ij}) \\ & + \mathbb{1}_j \ln(PICK_COUNT_{ij}) \delta_j + \mathbf{Z}_{ij} \boldsymbol{\theta} + \epsilon_{ij} \end{aligned} \quad (6)$$

Figure 6 Histogram of pick time multipliers for all pickers

This allows us to estimate a learning curve, similar to Figure 3a, for each worker. We find that indeed pickers exhibit a wide range of learning curves. Figure 7 displays the predicted pick time as a function of experience for a small sample of the pickers in the data (Similar plots are found in Gans et al. (2010) and Arlotto et al. (2013) of call-center workers).

Figure 7 Learning curves for a sample of pickers

Another way to conceptualize the heterogeneity in learning is to examine the pick time multipliers, similar to Figure 6, but recognizing that now these multipliers are a function of experience. To be more precise, previously we estimated the pick time multiplier as $\exp(\alpha_j)$, which is not dependent on experience. In the heterogeneous learning model (Equation 6), the pick time multiplier is calculated as $\exp(\alpha_j + \delta_j \ln(PICK_COUNT_{ij}))$ and thus depends on picker j 's experience at pick i .

We calculate the pick time multipliers first for $PICK_COUNT = 100$ and then for $PICK_COUNT = 36,000$. The results show that the dispersion of the estimated pick time multipliers decreases substantially with experience. There is a great deal of variance in picker performance across novices, but as pickers gain experience this variance diminishes.

We also note that a picker's performance as a novice is not particularly indicative of their relative performance as an expert. We calculate the correlation between pickers' pick time multiplier at $PICK_COUNT = 100$ and at $PICK_COUNT = 36,000$. We find this correlation to be -0.10 ($p=0.02$), suggesting that not only are fast novices not necessarily fast experts, but rather, the negative result suggests that there is a slight tendency for the slower novices to become the faster experts. We refer the reader to [Arlotto et al. \(2013\)](#) for further analysis of how managers can optimally monitor and retain workers with heterogeneous learning.

6. Simulation

The results from Section 5 indicate an opportunity to improve OFC productivity by incorporating factors such as bin density and picker experience into the pick assignment. We use Monte Carlo simulation to examine these possibilities and estimate the magnitude of any improvements.

6.1. Simulation Design

We create a simulated rectangular OFC with an area of A square feet containing B bins arranged in a grid with aisle, row, and bin spacing similar to our study company's OFC. A single trial of the simulation consists of a picker starting at the bottom left corner of the OFC, walking through the OFC to pick I items for a single order, and then returning to the starting point. Other key model parameters are the number of identical copies of a given SKU available in the OFC (C), and the experience of the picker (E) in units of $\ln(PICK_COUNT)$.

There are several random variates generated for each trial of the simulation. First, for each of the I items (SKUs) in the order, we generate the C bins where the desired SKUs can be found. We assume that items are equally likely to be in any bin and that no bin holds more than one copy of a given SKU. For each of the identified bins, we generate a bin load from a negative binomial distribution and convert this to bin density by dividing by the bin size (3.13 ft.). The distribution approximates the bin density distribution of the study company (Table 1).

The pick list (which item from which bin in what sequence) is made according to one of two heuristics. The Distance-Only (DO) heuristic uses a simple single-step, nearest neighbor distance minimization heuristic in which the picker is sent to whichever of the potential bins is closest to the picker’s current position.³ This is repeated until all I SKUs have been picked, at which point the picker returns to the starting point. Thus, for the first leg of the route, the heuristic picks the closest of the approximately $C \times I$ potential bins (While we assume I unique SKUs in each order, the actual number of potential bins could be less than $C \times I$ if some of the SKUs in the order are located in the same bin.). After the first item is picked, the heuristic then chooses from the remaining $C \times (I - 1)$ bins the bin closest to location of the first pick, and so on. This heuristic provides a baseline performance measurement against which to compare improvements.

The Distance-Density-Experience (DDE) heuristic is a similar single-step minimization heuristic, but it minimizes expected time rather than distance. For each potential bin, we compute the expected pick time using the estimated coefficients from Model 3 of Table 4 (Equation 5 excluding the $\beta_3 \ln(PICK_COUNT)$ term). The expected time estimate is the total time to both travel to the next bin and to search for the item in the bin. This estimate incorporates the distance to the bin, the density of the bin, and the experience level of the picker. The DDE heuristic sequences the picks by selecting the bin with the lowest expected pick time, which is not always the closest bin. Thus, compared to the DO heuristic, we expect the DDE heuristic to lead to lower average pick times and lower visited bin densities but higher average distances.

Once the route is planned, the simulation generates a random time realization for each leg of the journey. Regardless of which heuristic is used to plan the route, the time realization is based on Model 3 of Table 4. The time realization is generated from $time_i = \exp(\mathbf{x}_i\beta + \epsilon_i)$ where $\mathbf{x}_i\beta$ is the linear predictor portion of the regression model and ϵ_i is a draw from a logistic distribution with mean zero and standard deviation of $\frac{\pi\gamma}{\sqrt{3}}$.

Each simulation run consists of 10,000 simulated orders being picked according to both the DO and the DDE heuristics. Key outcome measures are order picking time, distance, and bin density. We report both mean values and mean percentage change between routes planned according to the two heuristics.

Because the results are dependent on the simulation parameters (A, I, C, E), we test a range of values for each of the model parameters (Table 5). Note that for clarity we hereafter refer to the three levels of Experience as *Low*, *Medium*, and *High*. We repeat the simulation for all possible combinations of the model parameters, a total of 300 scenarios.

³ Prior work has shown that nearest neighbor algorithms can be acceptable heuristics for “Traveling Salesman” type problems and are commonly used in practice (Johnson and McGeoch 1997, Gutin et al. 2002, Gu et al. 2007). We explore solving the Traveling Salesman Problem in Section 6.5.

Table 5 Simulation Parameter Values

Parameter	Levels
Warehouse Area (000's sq. ft.): A	12, 24, 36, 48, 96
Items per Order: I	1, 2, 3, 4, 5
Available Copies per Item: C	10, 50, 100, 200
Experience: E	6.2, 8.9, 10.3

6.2. Routing with Distance, Density, and Experience

We define the base case values of the simulation parameters as: $A = 12$, $I = 4$, $C = 100$, and $E = \textit{Medium}$. We select these as the base case because they are similar to a single zone of the study company OFC on which the regression model parameters are estimated.

Our first purpose for the simulation is to compare pick routes planned according to the DO and the DDE heuristics. Table 6 shows summary statistics from the base case scenario simulations, as well as the related two other experience level scenarios. Looking first at the DO heuristic results we see that mean order picking time decreases with experience, echoing the empirical results shown in Figure 3. We also see that mean distance and density are identical across experience levels, as expected, since the DO heuristic disregards density and experience when assigning picks.

Table 6 Simulation Summary Statistics

Experience	DO Heuristic			DDE Heuristic			Mean Percent Change		
	Low	Medium	High	Low	Medium	High	Low	Medium	High
Order Pick Time (sec.)	150.8 (0.30)	132.4 (0.27)	123.0 (0.24)	140.3 (0.27)	129.7 (0.26)	122.3 (0.24)	-6.8 (0.05)	-2.0 (0.03)	-0.5 (0.02)
Distance per Pick (ft.)	42.2 (0.16)	42.2 (0.16)	42.2 (0.16)	95.1 (0.34)	61.3 (0.23)	47.1 (0.17)	157 (1.4)	61 (0.8)	19 (0.5)
Density per Pick (items/ft.)	15.7 (0.03)	15.7 (0.03)	15.7 (0.03)	9.3 (0.01)	11.2 (0.02)	13.1 (0.02)	-39 (0.13)	-27 (0.15)	-15 (0.14)

Standard errors shown in parentheses
10,000 orders: $A = 12$, $I = 4$, $C = 100$

Comparing the DO results to the DDE results we see that the DDE heuristic leads to reduced order pick times, larger distances, and lower bin density of those bins from which an item is picked for all experience levels. The rightmost panel of Table 6 displays the mean percentage change for each order picked according to the DDE heuristic compared to the DO heuristic (note that this is different from the percent change of the means which could be calculated directly from the values in the left and center panels of Table 6). Using the DDE heuristic to plan pick lists leads to reductions in order pick time of 6.8% to 0.5% depending on experience. What is most remarkable about these time savings is that they are achieved despite large increases in required distance. For example, *Low* experience pickers travel 157% farther under DDE routing yet still achieve mean time reductions of 6.8%. This is due to the pickers being sent to bins that are 39% less dense, on average. Similarly, *Medium* and *High* experience pickers encounter larger distances and lower bin densities

relative to the DO heuristic, but the change is smaller because the deleterious effect of bin density diminishes with experience. Stated differently, as experience increases, there are fewer times when it is worthwhile for a picker to travel more than minimal distance to reach a lower density bin.

Figure 8 provides graphical results of one-way sensitivities of each of the main simulation parameters on the outputs of order time and distance. Each graph plots the mean percentage change between the DO and DDE heuristic for 10,000 shipments of the given scenario. Figures 8a and 8b show the impact of experience on the outcome measures, just as shown in the right panel of Table 6.

Figures 8c and 8d show that the time savings benefit of the DDE heuristic diminishes with larger warehouses (holding I , C , and E constant). This happens because as the warehouse gets larger (i.e., has more bins of same size, spacing, and density distribution), the expected distance between copies of an item grows and it becomes less likely that the reduced search time due to a low density bin will be sufficient to warrant the extra travel.

Figures 8e and 8f show that time savings and distance traveled increase with the number of copies per item in the OFC. This result occurs because with higher copies per item, each item gets more draws from the density distribution and thus there is a higher probability of having bins with low densities worth traveling extra distance to reach.

Lastly, Figures 8g and 8h show that, similar to copies per item, time savings and distance increase with increased items per order. This result occurs for similar reasons to the copies per item result. When planning the pick route, the DDE heuristic chooses the first pick from the $C \times I$ potential bins. As I (or C) increases, there are more potential bins for the heuristic to choose from and thus a higher probability of finding a bin with a performance-improving distance and density combination.

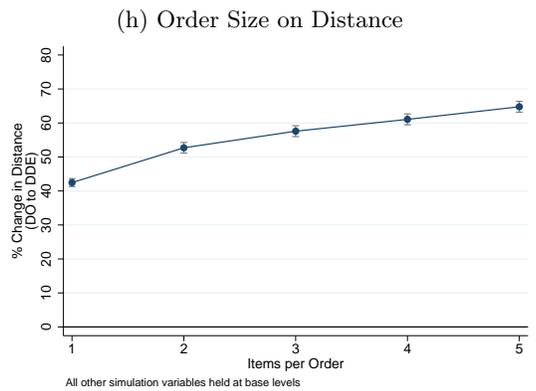
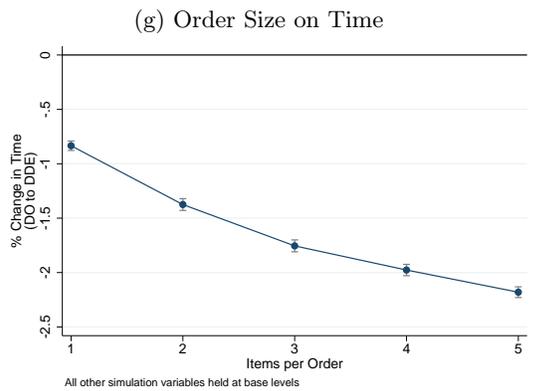
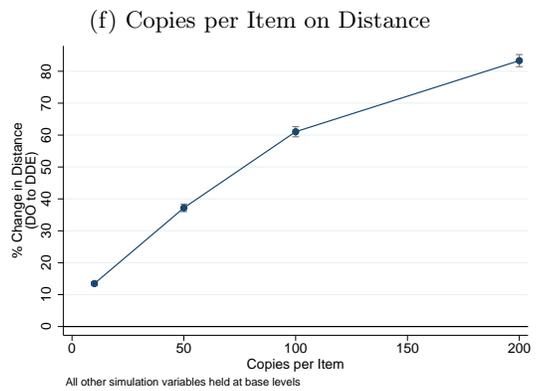
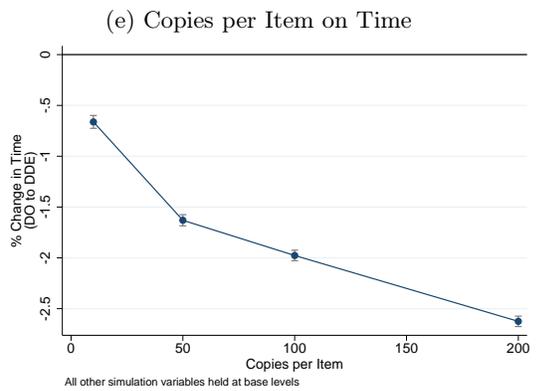
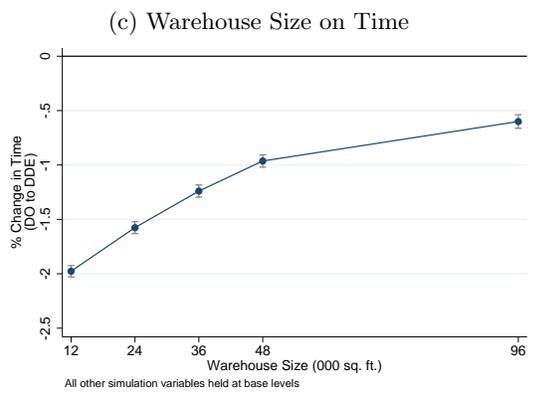
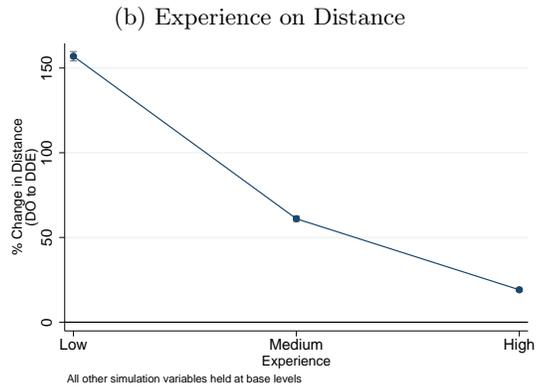
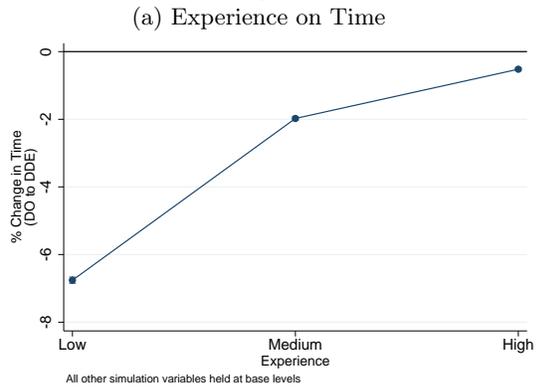
In summary, the above simulation results show that changing from a distance-only pick planning heuristic to a time-based heuristic which incorporates distance, density, and experience leads to improvements in order picking times. Inexperienced workers exhibit the biggest improvements. We further observe that these improvements are largest for small warehouses or when there are many copies of an item, or both.

6.3. Task Assignment

Having shown that taking distance, density, and experience into consideration is an effective way to improve average pick times of any given picker, we now consider the problem of incorporating distance, density, and experience into assigning orders across a cohort of heterogenous pickers.

We expect that in the presence of heterogenous pickers, OFC performance can be improved beyond that of simply randomly allocating orders across pickers. To estimate the magnitude of this improvement we compare the simulated performance of a cohort of workers picking a fixed number of orders under three different assignment scenarios. For this analysis we continue to focus on the

Figure 8 Impact of DDE vs. DO routing on pick time and distance



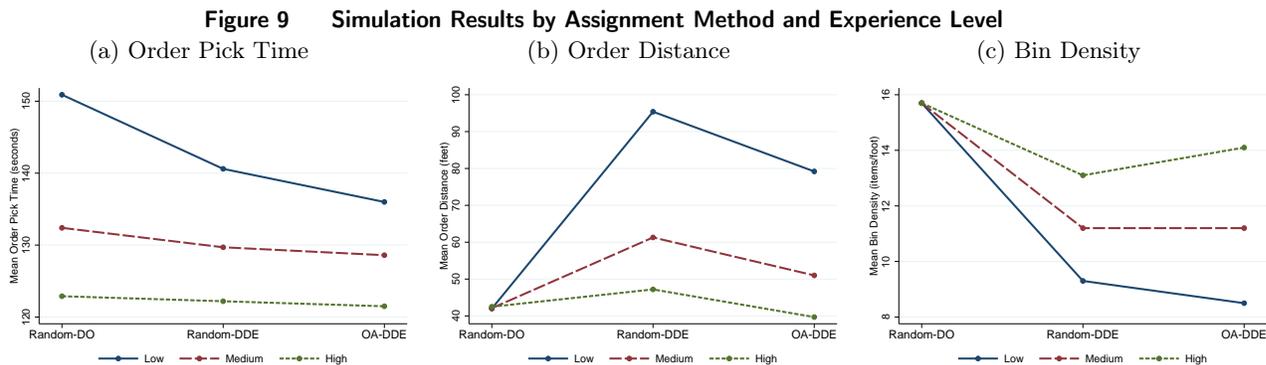
Note: Error bars indicate 95% confidence interval

base case OFC as defined in Section 6.2.⁴ Further, we assume one worker of each experience level, *Low*, *Medium*, and *High* are available to pick orders. As in Section 6.1, we generate 10,000 random orders to be picked.

In the Random-DO scenario, the 10,000 orders are assigned to the pickers at random and the pick routing is done according to the DO heuristic of Section 6.1. In the Random-DDE scenario, the orders are assigned following the same random assignment as the Random-DO scenario, but pick routes follow the DDE heuristic.

The third scenario is the Optimal Allocation-DDE (OA-DDE) scenario. In this scenario we treat the assignment of orders to pickers as an assignment problem with the objective of minimizing the total expected labor time necessary to pick all 10,000 orders, subject to the constraint that each worker be assigned the same number of orders to pick (Winston and VenKataramanan 2003).⁵ We use the DDE heuristic to generate an expected pick time for every picker-order dyad, and then use the Hungarian Algorithm (Kuhn 1955) to optimally assign orders across the three pickers.

For each of the three scenarios, we compute the mean time, distance, and bin density per order. The results from the three scenarios are presented in Figure 9. Note that the Random-DO and Random-DDE results are similar to those presented in Table 6 because those results also assume random assignment of orders to pickers. We include these results again in Figure 9 for ease of comparison to the OA-DDE scenario results.



There is a large pick time reduction obtained by switching from the Random-DO to the Random-DDE scenario (Figure 9a). However, this time improvement is obtained at the expense of pickers of all experience levels walking longer distances (Figure 9b). For example, for the *Low* experienced

⁴ The results presented in this section are robust to different sizes of OFC, copies of items and items per order.

⁵ The “equal number of orders” constraint prevents the algorithm from assigning all the orders to the fastest (most experienced) worker. The improvement gained by using the OA-DDE system increases if the constraint is relaxed.

picker, the mean order picking time is reduced by 7% (from 151 to 141 seconds) while the mean distance increases 126% (from 42 to 95 feet).

Under the OA-DDE scenario we observe an additional time savings across all experience levels, albeit smaller in magnitude than the change from Random-DO to Random-DDE. The time reductions between the Random-DDE and the OA-DDE scenario are 3.2%, 0.8%, and 0.5% for the *Low*, *Medium*, and *High* experience levels, respectively.

An additional benefit of the OA-DDE scenario is that there is a reduction in the mean distance per order relative to the Random-DDE scenario. This is particularly beneficial because the time improvement obtained with the Random-DDE scenario was at the expense of more walking by the pickers. The distance reductions between the Random-DDE and the OA-DDE scenario are 17.0%, 16.7%, and 15.9% respectively for the *Low*, *Medium*, and *High* experience levels. We note that the *High* experience worker actually travels less per order under OA-DDE than either of the other two scenarios. This is because the *High* experience worker is comparatively faster at picking high density picks regardless of distance. Thus the optimal assignment gives lower density but farther picks to the other workers and sends the *High* experience worker to the closer, higher density picks.

Figure 9c displays the mean bin density for items picked and helps explain the above results. Switching from the Random-DO assignment to Random-DDE assignment leads to reduced bin densities as pickers of all experience levels are routed to lower density bins to achieve faster pick times. When OA-DDE assignment is used, the *Low* experience workers (who are most sensitive to density) get a disproportionate share of the “easy” (low density) assignments, while the *High* experience workers take on “harder” (high density) picks.

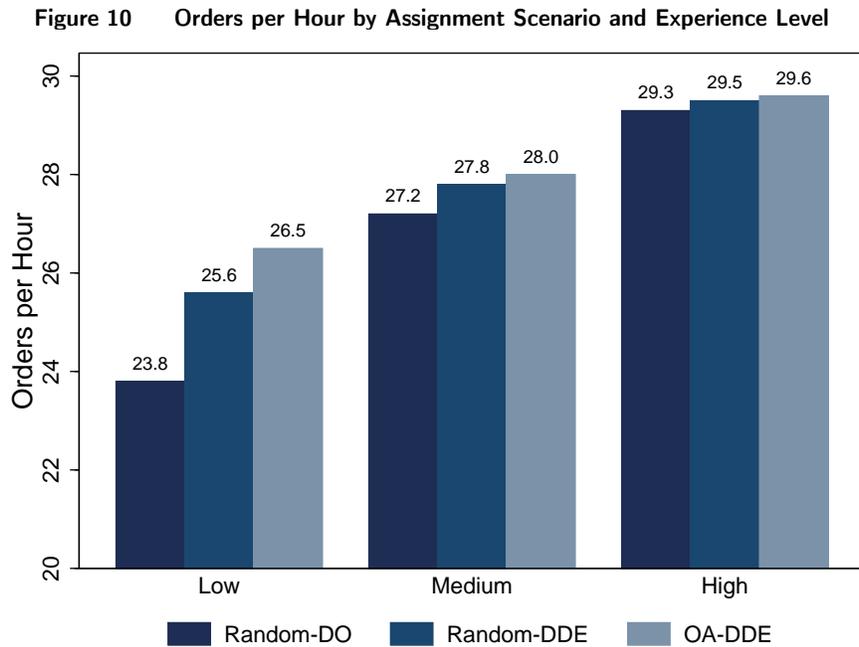
An alternative metric to compare the three assignment scenarios is the mean number of orders per hour that each type of picker completes (Figure 10). This metric provides a simple productivity measure. We see that *Low* experience pickers complete almost 3 more orders per hour under OA-DDE as compared to Random-DO. Improvements are smaller in magnitude for more experienced pickers.

If we consider one picker of each experience level working for one hour, on average they complete 80.3 orders ($23.8 + 27.2 + 29.3$) under the Random-DO scenario, 82.8 orders ($25.6 + 27.8 + 29.5$) under the Random-DDE scenario, and 84.1 ($26.5 + 28.0 + 29.6$) under the OA-DDE scenario. This represents a 4.7% increase in workforce productivity from the baseline Random-DO routing and assignment method to the OA-DDE method.

6.4. Effects of Bin Size⁶

In the above analysis (Sections 6.2 and 6.3), bin size and the parameters of the random distribution of items per bin are held constant at values similar to the study site, while several values of the

⁶ We thank an anonymous referee for the suggestion of this analysis.



other model parameters are tested. However, it is possible that picker productivity can be further improved by optimally setting the bin size and thus the density, as well.

We repeat the DDE heuristic simulations of Sections 6.2 under five levels of bin size: 2 ft., 2.5 ft., 3 ft. (same as Section 6.2), 3.5 ft., and 4 ft. We hold the distribution of items per bin fixed so that increasing the bin size (feet of hanging rod) reduces bin density (items per foot). However, changing the bin size increases the average distance between bins and necessitates a change in OFC size as well. For simplicity, we keep the layout of the OFC the same (e.g., same number of rows and cross aisles, same aisle widths, etc.) but allow the length of the rows and thus the width of the simulated building to change as the bin size changes. When comparing performance across the bin size parameters, we keep all other random variates the same. In other words, the randomly generated number of items in each bin, the locations of each item, and the SKUs in each order are the same across each simulation run.

The results show that lowering bin density by increasing bin size reduces the mean picking time. Just as in Sections 6.2, the magnitude of the effect varies with experience (E) and the number of copies (C), but is on the order of a 10% for low experience pickers and 3% for high experience pickers as bin width varies from two feet to four feet (results table not shown due to space constraints).

What is particularly interesting about these results is that not only does mean pick time decrease with increased bin size, but mean distance also decreases. One might expect distance to increase as bins get larger and farther apart. In contrast, the simulation shows that the reduced density of the larger bins has a large enough impact on expected pick time to result in sending the picker to a

different, closer bin, even though it has a higher density than a farther away bin. Stated differently, as density decreases, distance becomes the bigger driver of routing.

Our results, however, do not demonstrate that there is an “optimal” bin density. Rather, this analysis highlights the complexity of determining an optimal density. Bin density, number of copies of an item, number of SKUs, size of the OFC, cost of the OFC, OFC layout, and workforce experience are all somewhat interconnected and must be jointly considered when determining an optimal bin size. While this is an interesting avenue for future research, optimizing across all of these factors is beyond the scope of this paper.

6.5. Robustness: Optimal Routing via Generalized TSP

We conclude this section by comparing exact optimal routing to the nearest-neighbor heuristic used in the analysis above in Sections 6.2, 6.3, and 6.4. In our setting, solving the Traveling Salesman Problem (TSP) to find the optimal routing solution presents a challenge. The picker receives a list of I items to be picked. Each one of these items have C identical copies in the OFC. Hence, this is a generalization of the standard TSP, known as the a “Set TSP” or “Generalized TSP” (GTSP), as it requires finding a shortest tour that visits all specified subsets (I) (Srivastava et al. 1969). The standard TSP is a special case of the GTSP when all subsets to be visited are singletons (e.g., when $C = 1$). The GTSP is NP-hard, and thus it is only possible to solve problems of limited size.

One of our contributions is showing that searching for an item in a bin is time consuming and this has been previously ignored as a relevant factor in the routing algorithms. In other words, our contribution is in proposing a new objective function based on distance, density, and experience, rather than distance alone, not a new optimization algorithm or heuristic.

In Section 6.2, we used the nearest-neighbor heuristic to show how picking performance can be improved by taking distance, density, and experience into account. We now repeat the analysis using the GTSP. We solve the GTSP by full enumeration for some reasonable size problems and show that the improvement in performance of the DDE versus the DO objective is similar to the improvement when using the nearest-neighbor heuristic.

Similar to the analysis in Section 6.2, we consider a warehouse of 12,000 square feet ($A=12$), five items per order ($I=5$), and three levels of experience ($E=6.2, 8.9, 10.3$). However we limit the number of available copies (C) to five to keep the scale of the problem small enough to solve in a reasonable amount of time. Following the same methodology as before, we first perform routing optimization via the nearest neighbor heuristic under the DO and the DDE objective, and then we repeat the analysis by solving the GTSP optimization under both the DO and the DDE objective. For the GTSP optimization we solve 3,125 (5^5) TSP routing problems that correspond to five items at five different locations and select the one that gives the minimum distance or expected pick time, respectively.

As expected, the GTSP full enumeration optimization approach performs better than the nearest-neighbor heuristic both under the DO and the DDE objective, and the improvement is consistent across the three levels of experience.⁷ However, more relevant for our purposes, the improvement obtained by using the DDE objective rather than the DO objective is similar in magnitude regardless of the optimization method used (3% for low experience, 1% for high experience). In fact, the improvement from using DDE rather than DO is slightly larger under GTSP optimization than under the nearest-neighbor heuristic.

In practice, solving the GTSP exactly may not be practical. For this small-scale problem (5 items, 5 copies), the nearest-neighbor heuristic requires 2.5 seconds, while the GTSP full enumeration algorithm requires 116 seconds, on average. Solving a larger scale problem, for example, a shipment of 10 items with 10, 50, or 100 copies of each item (realistic numbers for the retail setting), quickly becomes infeasible as the GTSP problem explodes (i.e., solving 100^{10} individual TSPs).

While we have estimated the benefits of the DDE objective function via full enumeration of a small GTSP problem and via the nearest neighbor heuristic for larger problems, there exist several other heuristics for optimizing the GTSP (e.g., Noon and Bean 1993, Yang et al. 2008). We anticipate that the gains from using the DDE objective rather than the DO objective are similar regardless of the optimization heuristic used. We note that some of the existing GTSP heuristics require that the expected costs of all arcs in the network be known *ex-ante*. This may at first appear to rule out the use of the DDE objective due to the fact that it is based on the cumulative experience of the picker at the time of each pick, which is itself a function of the sequence of the optimized pick list.⁸ However, one can transform the experience-dependent problem into a standard GTSP by creating “dummy” copies of each node which represent the item being picked with different levels of experience. Thus, there need to be as many dummy nodes for each item as there are experience levels the picker could have when picking the given item. The arc costs are then constructed to ensure that the chosen path always moves between nodes with sequentially increasing experience. Thus, any heuristic that can be used with the DO objective can also be used with the DDE objective, albeit with a larger network of nodes.

It is possible that for some situations, solving the experience-dependent GTSP via dummy nodes becomes impractical due to the problem size. In such situations, optimizing the DDE objective via the nearest-neighbor heuristic or via a fixed-experience GTSP (experience is updated once per route) are both good alternatives and will perform at least as well, if not better than optimizing on distance only.

⁷ Due to space constraints we do not present the full table of results. They are available from the authors by request.

⁸ We thank an anonymous referee for pointing out this challenge.

7. Discussion and Conclusion

The rise in multichannel retailing has pushed retailers to broaden their view of warehouse operations from shipping cases to stores to shipping single items to individual customers. These two activities are substantively different. One common practice in OFC design is chaotic storage due to its efficient use of space. However, one drawback of chaotic storage is that because dissimilar items share a common location or bin, a new step is added in the picking process: searching.

This paper shows that pick time, which incorporates both travel time and search time, is a function of the minimum required travel distance, the number of items in the bin, and the density of the items within the bin. When pickers encounter bins with many items, they need to search through more items, on average, before finding the right item, leading to longer pick times. Likewise, high bin density can impede the search by making it more difficult to see or access the items.

We also show that pick times improve as workers gain picking experience. Further, by interacting experience with distance and density we find that the performance gains from learning happen through a reduction in the marginal effect of density but not distance. In other words, the deleterious effect of density diminishes as pickers become more experienced, but the effect of each foot of travel distance remains unchanged.

Having spent time in the OFC both observing and picking items, we can offer a few conjectures on the mechanisms behind these results. First, the layout of the OFC and the numbering of the bins is simple and intuitive with rows of hanging rods divided into consecutively numbered bins. Further, the signage throughout the OFC is plentiful and clear. Thus, it is reasonable that pickers become proficient at navigating the OFC quite quickly and show no significant improvement with experience.

Regarding the learning that occurs in the searching process, we can envisage several possible mechanisms related to improved use of information. The pickers have three key pieces of information that they can use to help identify the correct item in the bin. First is the picture of the item which appears on their computer screen. This provides visual clues such as color and garment type (e.g., blouse, pants). The second piece of information is the text description of the item which appears on both the pick list and the computer (e.g., “Black S Sanctum Fizzo Skinny Ponte Pants,” “Purple L 52Holly Moni Stud Detailed 3/4 Sleeve Blouse.”). These descriptions provide the color, size, brand, model and type. The third piece of information is the six to eight digit item identification number, which is akin to a serial number and uniquely identifies the garment. The item ID is provided to the picker on the pick list and the picker’s computer. On the garment, the item ID and related barcode are printed on a small tag which is attached to the garment.

We believe learning can happen across all three types of information. With practice, pickers get better at looking for visual clues, such as color and pattern. Also, with experience, workers become

more familiar with the descriptive text which helps identify items (we certainly had to learn what “ponte pants” and “maxi skirts” are). Lastly, workers learn where the item ID tags are commonly placed on different types of garments, and perhaps develop mental routines that help them remember or compare the multi-digit ID numbers (Kliegl et al. 1986). It is also possible that pickers become familiar with specific SKUs over time, but given that the study company is largely a fast-fashion company with rapid product turnover (similar to Caro and Gallien (2007)), this knowledge would quickly become obsolete and need to be frequently refreshed, rather than improving continuously over time.

Our proposed DDE heuristic improves pick time performance by trading off distance and density. This is particularly true for novice pickers, and under DDE they systematically visit lower density bins. This could raise a concern about impeded learning due to having predominantly “easy” picks. We test for differential learning effects of low and high density picks and do not find support for such an effect. There is also a possible concern with fatigue affects over the course of the day. We test for this mechanism but do not find support for it. Further details are available from the authors.

We use simulation to show that pick times can be improved by incorporating density and experience into the pick routing. This happens by minimizing expected pick time rather than distance. It is interesting to note that this performance improvement comes from making use of both the specifics of the task and of the worker. For example, it is not sufficient for the manager to know that SKU 123 needs to be picked. To fully understand the task, the manager needs to know that at this moment SKU 123 is available in 100 different locations in bins with 100 different densities. Still, that is not sufficient. To plan the pick list, the manager needs to know the experience level of the person who will pick the item. It is through properly combining the details of the task and the worker that the best performance is achieved.

This motivates our analysis of the problem of how to assign tasks across a heterogenous workforce. We show that by taking task and worker specifics into consideration, productivity can be improved further. In addition, the optimal assignment has the added benefit of reducing the mean distance traveled by workers.

Recently, there have been several articles in the media highlighting the demanding work conditions at OFCs (e.g., McClelland 2012, Jamieson 2015). These articles emphasize how uncomfortable the conditions can be at OFCs (e.g., extreme heat or cold), and how the picking task itself can be physically demanding and even present health risks (McClelland 2012, Jamieson 2015). Furthermore, when describing the picking task, pickers frequently mention the mental stress and illnesses that can result from the work (McClelland 2011, Dixon 2013). Such problems can be exacerbated in chaotic storage systems where pickers commonly are required to locate an item within a fixed period of time. If an item is missing from the bin, the system can require the picker to scan every item in the

bin to “prove” that the item they cannot find is truly missing (McClelland 2012). Not surprisingly, OFC jobs tend to have high employee turnover (Min 2007).

However, our results demonstrate the benefit of low turnover and high worker experience in the OFC. Experienced pickers are significantly more productive than inexperienced ones (Figure 10), and this additional productivity translates into lower operational costs for the retailer. Or, conversely, retailers could consider creating financial incentives aimed at reducing turnover of experienced pickers. The implementation of our suggested approach should be done carefully since it can raise a concern related to the fairness of the assignment. We believe that transferring part of the benefits from the efficiency improvement to the picker (e.g., higher wages) can be a good way to mitigate the fairness issue.

To support this suggestion we present the following calculations that illustrate the incentive opportunity that the retailer could offer. Consider High, Medium, and Low experience pickers (as defined in Section 6.1). Our analysis estimates mean pick times of 41.3, 45.0, and 53.3 seconds respectively, which convert to picks per hour of 87, 80, and 68, respectively. If all pickers are paid \$9 per hour (California minimum wage at the time of the data (California 2017)), the company is effectively paying \$13.24 per 100 picks from a low experience worker, but only \$11.25 and \$10.34 per 100 picks from medium and high experienced workers, respectively. Paying medium and high experience workers \$10.62 and \$11.52 per hour, respectively, would bring these workers to parity in terms of pay for productivity. Alternatively, the company could simply pay for performance (i.e., pay per pick). The company could also consider transferring the efficiency gains from more experienced workers to working condition improvements that will increase retentions (e.g., air conditioning, more frequent breaks, nice break rooms, etc).

Retailing in general, and online retailing in particular, must deal with high demand seasonality. In 2014, 23% of U.S. annual online sales happened during the holiday season, and retailers collectively added over 700,000 seasonal employees to handle the demand (Statista 2015, Allen 2015).⁹ Amazon.com alone hired 80,000 seasonal workers for its OFCs in 2014 and 100,000 in 2015 (Satariano and Soper 2015). Our analysis provides insight into how to make best use of the large number of part time and inexperienced workers that are hired during the peak season. We show that optimally assigning orders across the available pickers increases productivity of the OFC while allowing the inexperienced workers to learn and gain experience on shorter, lower density pick assignments.

Finally, we believe that retailer fulfillment capabilities will increasingly become a strategic component of retailers’ operations. Today, many retailers rely on third party fulfillment companies to handle their online order fulfillment needs (Soshkin 2015). This practice is reasonable when

⁹ The National Retail Federation defines the “holiday season” as the 61 days of November and December.

online sales represent only a small portion of the retailer's business. However, as the online channel grows and represents a larger portion of sales, and as customers increasingly demand a seamless, multichannel shopping experience, retailers will face increased pressure to make order fulfillment a key strategic part of the business rather than something that can be outsourced to a third party. Our work can help companies get the most out of their order fulfillment centers.

Appendix

A. Density and Visual Search ¹⁰

Visual search is a fascinating and active research field. Although it is true that finding a particular item becomes easier when the other items near by are more diverse, there is not yet a proven method or metric to quantify the complexity of a visual search. We present a brief review of several relevant papers.

Bravo and Farid (2008) develop a “clutter measure” and show how item search time varies with the level of clutter in an image. The weakness of this measure for our purposes is that the measure assumes that the target item must be of uncertain color, orientation, and simple features (e.g., the searcher is told to “find the hat”). It is clear that the problem faced by pick workers is not consistent with these assumptions, since they know exactly what item they are looking for. Further, the images must be unstructured and target items must be drawn from a pool of similar items (i.e., not be particularly salient). Thus, the assumption rules out the very visual salience that we would like to measure and take advantage of.

There is a large amount of research studying how to define salience and different ways to measure it (see for example Rosenholtz et al. (2007)). Nothdurft (2002) presents a number of results related to the importance of the “distinctiveness” of a target from nearby non-targets. According to Nothdurft, the main factors driving the salience of an item are, luminance, color, motion, orientation, or depth. The author argues that targets with high feature contrast appear as salient. Salient targets are found quickly, regardless of image set size (number of other objects in the search). Search for non-salient targets is slower and tends to be affected by set size, which would be consistent with our finding that pick time increases with bin density.

Similarly, Bravo and Farid (2004) find that search time goes up with distractors (i.e., other items in the image), and that the effect is exacerbated by items of compound color. The authors also assume that the target is not salient. Other related papers which discuss the issue of target-distractor interaction in the search process include Rosenholtz et al. (2005), Rosenholtz et al. (2007), and Duncan and Humphreys (1989). In short, the science of visual search has not advanced yet to the point of having a good salience measure that we can use for our situation.

Although the current literature does not offer a metric to capture “product variety in a bin” or “target item salience,” we performed the following study to get a sense of the magnitude of the effect of the visual complexity of the search and whether bin density remains a valid predictor once visual complexity is controlled for. We designed and implemented a computer-based lab experiment in which participants search for a single item in a “rack” of clothes, similar to what the pickers do at the warehouse. The following paragraphs describe this experiment.

From a set of pictures of 100 individual items or garments (25 each of red, blue, green, and floral color/pattern), we identified 16 “target” items for which the subjects would search (four from each color). We generated 48 sets or “racks” of clothing with between 10 and 20 individual items per rack. The racks were generated such that each rack contained 3 target items, plus another 7 to 17 items randomly drawn from the pool of items, and each target appears in 7 to 12 racks. With three targets per rack, this gives a total of 144 rack-target search tasks. The images below show a sample of the actual racks used in the experiment:

¹⁰ We thank an anonymous referee for the suggestion of this analysis.

Figure 11 Sample of Actual Racks

Each participant in the experiment is asked to complete a random selection of 50 (out of the 144) individual search tasks. Each individual task is presented to the participant in the following fashion. After a description of the task, the first screen shows the participant the item she needs to find:

Figure 12 Sample Item to be Found

The next screen shows the rack to search. The participants are asked to click on the target item on the rack. (The subject also has the option to go back to look at the target item again).

When the subject clicks on the correct item they move on to see the next item to be found. If the subject clicks on the wrong item they receive an error message asking them to try again (they can not proceed until the correct item is clicked). By design the target item is never the leftmost item on the rack. Thus, the target item is always partially covered.

The study was completed by 250 participants during a two day period between March 15 and March 16, 2017. The participants were recruited through the Amazon Mechanical Turk platform and received a payment of 50 cents of a dollar to complete the task. Considering the average time that the task took to

Figure 13 Sample Rack to Search.

complete and the payment we offered, the effective hourly rate for the task was significantly higher than an average task on the platform. This allowed us to recruit workers that have the highest qualifications by filtering only workers with the highest evaluations through the platform system. Given that the workers' personal information was not critical to the task we assigned, we did not collect any personal information on the workers.

After completing their tasks, the 250 participants provided 12,500 observations ($250 \cdot 50$). For each observation, we record the time to complete the task (duration from first view of the target to correctly clicking on the target), the number of times that the participant clicked on the "view item again" button, and the number of times that the participant clicked on a wrong item.

We exclude 26 observations with task time of greater than 60 seconds. Of the remaining 12,474 observations, the average task time is 5.4 seconds with a standard deviation of 5.4 seconds. The median time is 3.7 seconds and the 95th percentile is 14.5 seconds.

We want to see if the impact of number of items in the rack (*RackCount*) is a significant predictor of task time even after controlling for the complexity of the pick. To do this we estimate the following model:

$$\ln(\text{TIME}_{ij}) = \alpha_j + \beta_1 \text{RackCount}_i + \mathbf{Picker}_j \beta_2 + \mathbf{Task}_i \beta_3 + \epsilon_{ij} \quad (7)$$

\mathbf{Task}_i is a vector of dummy variables for each one of the 144 item-rack search task combinations. This controls for the inherent difficulty of finding the given item in the given rack, which is due to the visual salience of the item relative to the rest of the rack. \mathbf{Picker}_j is a vector of dummy variables that correspond to each participant. *RackCount*_{*i*} is the number of items in an individual rack image. Because the rack images are all of the same size or length, *RackCount* is a direct analog of the *DENSITY* variable in the analysis in Section 5.

Table 7 shows the results of this analysis. Column 1 presents the results when we only include the \mathbf{Picker}_j fixed effect variables. Column 2 shows the results when we add the *RackCount*_{*i*} variable. We see that it is positive and significant indicating that the task takes longer for racks with more items. This corresponds to and corroborates the density effect we show in Section 5.2 of the paper.

Column 3 shows the results of the complete model including the \mathbf{Task}_i fixed effect for each search task. These results provide two insights. First, we see that the R-squared value jumps from 0.205 to 0.488 from

Table 7 Density Analysis

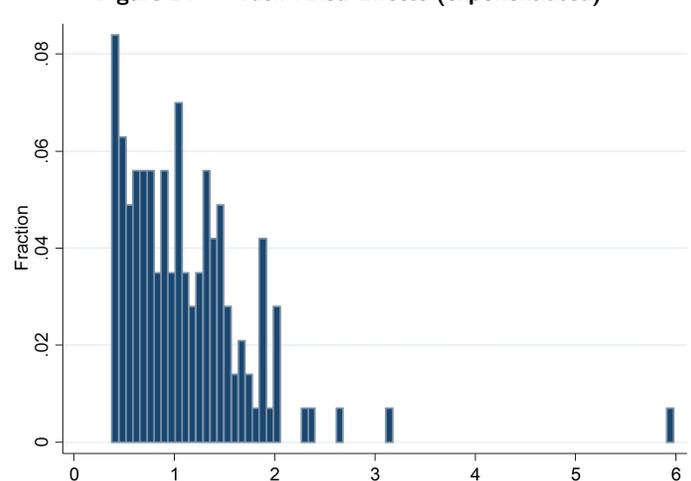
	1	2	3
<i>RackCount</i>		0.030*** (0.000)	0.143*** (0.000)
Picker FE	Yes	Yes	Yes
Task FE	No	No	Yes
N	12,474	12,474	12,474
R^2	0.179	0.205	0.488

Robust standard errors in parentheses

* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Columns 2 to 3 suggesting that the visual characteristics of the task have a meaningful impact on the search time and can explain a great deal of the variation in pick time. Second, we see that even after controlling for the complexity of the search task, the number of items in the rack still has a positive and significant effect on the search time, indicating that the piece of information would still be a valuable predictor of pick time even if we had a measure of visual variety or salience.

To get a sense of the magnitude of the fixed effects of the 144 search tasks, we exponentiate the Task coefficients to convert them into time multipliers, relative to a task of average difficulty (the omitted variable in the set of Task dummy variables is intentionally selected to be an average difficulty task). Figure 14 presents these exponentiated coefficients. We see that these multipliers range from about 0.2 to 6, with the bulk of them being between 0.2 and 2. This means that “easy” search tasks have an average search time of about 20% of the average (1.1 seconds) while “harder” tasks require up to about twice as much time as the average task (10.8 seconds).

Figure 14 Task Fixed Effects (exponentiated)

To summarize, both the literature on visual search and our lab experiment support that the difficulty of a visual search task is a combination of the number of objects in the search (how many items on the rack), the variety of the objects (colors, styles, patterns), and the specific target item (the desired piece of clothing). Of these three factors, given the current state of visual search science, only the first can be quantified and

operationalized to be included in an algorithm that predicts search times. The results of our lab experiment replicate the finding from the main paper that this factor, item count (related to density), is statistically meaningful to establish the search time of an item in the rack and hence can be used as a proxy for the complexity of the task.

Acknowledgments

The authors gratefully acknowledge the managers from the anonymous company for graciously sharing their time, data, and operational expertise. This paper could not have been written without their support. The authors also gratefully acknowledge helpful comments from James G. Morris, Kenneth R. Baker, and Hany Farid, as well as participants at the 2016 Behavioral Operations Workshop, and the 2017 Consortium for Operational Excellence in Retail (COER) Conference. The authors also thank Robert Burnham for his invaluable help in developing the simulations tools. Finally, the authors also thank seminar participants at The Wharton School for their insights.

References

- Allen KG (2015) Holiday trends and expectations: NRF's 2015 holiday survival kit. online, URL https://nrf.com/sites/default/files/2015%20NRF%20HSK_102015_Final.pdf.
- Arlotto A, Chick SE, Gans N (2013) Optimal hiring and retention policies for heterogeneous workers who learn. *Management Science* 60(1):110–129.
- Author (2016) Personal communication with managers.
- Banjo S (2013) Rampant returns plague e-retailers. *The Wall Street Journal* URL <https://www.wsj.com/articles/rampant-returns-plague-eretailers-1387752786>.
- Bartholdi JJ, Hackman ST (2014) *Warehouse and Distribution Science* (The Supply Chain and Logistics Institute: Georgia Institute of Technology), URL www.warehouse-science.com, release 0.96.
- Batt RJ, Terwiesch C (2015) Waiting patiently: An empirical study of queue abandonment in an emergency department. *Management Science* 61(1):39–59.
- Becker-Peth M, Katok E, Thonemann UW (2013) Designing buyback contracts for irrational but predictable newsvendors. *Management Science* 59(8):1800–1816.
- Beer R, Ahn HS, Leider S (2014) Can trustworthiness in a supply chain be signaled? *Ross School of Business Paper* (1251).
- Bell DR, Gallino S, Moreno A (2014) How to win in an omnichannel world. *MIT Sloan Management Review* 56(1):45.
- Bendoly E, Donohue K, Schultz KL (2006) Behavior in operations management: Assessing recent findings and revisiting old assumptions. *Journal of Operations Management* 24(6):737 – 752, ISSN 0272-6963.
- Boudreau J, Hopp W, McClain JO, Thomas LJ (2003) On the interface between operations and human resources management. *Manufacturing & Service Operations Management* 5(3):179–202.

- Bozer YA, White JA (1990) Design and performance models for end-of-aisle order picking systems. *Management Science* 36(7):852–866.
- Bravo MJ, Farid H (2004) Search for a category target in clutter. *Perception* 33(6):643–652.
- Bravo MJ, Farid H (2008) A scale invariant measure of clutter. *Journal of Vision* 8(1):23–23.
- California So (2017) History of california minimum wage. URL <https://www.dir.ca.gov/iwc/MinimumWageHistory.htm>.
- Caro F, Gallien J (2007) Dynamic assortment with demand learning for seasonal consumer goods. *Management Science* 53(2):276–292.
- Chao L (2015) The rising cost of free returns. *The Wall Street Journal* URL <https://www.wsj.com/articles/the-rising-cost-of-free-returns-1444398695>.
- Cleves MA, Gould WW, Gutierrez RG, Marchenko Y (2010) *An introduction to survival analysis using Stata* (Stata Press).
- Coyle JJ, Bardi EJ, Langley CJ (1996) *The Management of Business Logistics* (St. Paul, MN: West Publishing).
- Davis AM, Katok E, Santamaría N (2014) Push, pull, or both? a behavioral study of how the allocation of inventory risk affects channel efficiency. *Management Science* 60(11):2666–2683.
- de Koster R, Le-Duc T, Roodbergen KJ (2007) Design and control of warehouse order picking: A literature review. *European Journal of Operational Research* 182(2):481 – 501.
- Dixon H (2013) Amazon workers “face increased risk of mental and physical illness”. *The Telegraph* URL <http://www.telegraph.co.uk/technology/amazon/10472094/Amazon-workers-face-increased-risk-of-mental-and-physical-illness.html>.
- Duncan J, Humphreys GW (1989) Visual search and stimulus similarity. *Psychological review* 96(3):433.
- Dutton JM, Thomas A (1984) Treating progress functions as a managerial opportunity. *Academy of management review* 9(2):235–247.
- Fisher M, Gallino S, Xu J (2016) The value of rapid delivery in online retailing. *Working Paper* .
- Foley R, Frazelle E (1991) Analytical results for miniload throughput and the distribution of dual command travel time. *IIE TRANSACTIONS* 23(3):273–281.
- Frei FX, Ely RJ, Winig L (2011) *Zappos.com 2009: Clothing, Customer Service, and Company Culture* (Harvard Business School Publishing).
- Gallino S, Moreno A (2014) Integration of online and offline channels in retail: The impact of sharing reliable inventory availability information. *Management Science* 60(6):1434–1451.
- Gans N, Liu N, Mandelbaum A, Shen H, Ye H, et al. (2010) Service times in call centers: Agent heterogeneity and learning with some operational consequences. *Borrowing Strength: Theory Powering Applications—A Festschrift for Lawrence D. Brown*, 99–123 (Institute of Mathematical Statistics).

- Gao F, Su X (2016) Omnichannel retail operations with buy-online-and-pick-up-in-store. *Management Science* Forthcoming.
- Gibson DR, Sharp GP (1991) *Planning and Control of Material Handling Systems*, chapter Initial System Configuration for Miniload Automated Storage/Retrieval Systems, 51–66 (American Society of Mechanical Engineers).
- Goetschalckx M, Ratliff HD (1990) Shared storage policies based on the duration stay of unit loads. *Management Science* 36(9):1120–1132.
- Greene WH (2012) *Econometric Analysis* (Prentice Hall), 7th edition edition.
- Gu J, Goetschalckx M, McGinnis LF (2007) Research on warehouse operation: A comprehensive review. *European Journal of Operational Research* 177(1):1–21.
- Gu J, Goetschalckx M, McGinnis LF (2010) Research on warehouse design and performance evaluation: A comprehensive review. *European Journal of Operational Research* 203(3):539 – 549.
- Gudehus T, Kotzab H (2012) *Comprehensive logistics* (Springer Springer Science & Business Media).
- Gutin G, Yeo A, Zverovich A (2002) Traveling salesman should not be greedy: domination analysis of greedy-type heuristics for the TSP. *Discrete Applied Mathematics* 117(1):81–86.
- Hausman WH, Schwarz LB, Graves SC (1976) Optimal storage assignment in automatic warehousing systems. *Management Science* 22(6):629–638.
- Huckman RS, Staats BR, Upton DM (2009) Team familiarity, role experience, and performance: Evidence from indian software services. *Management science* 55(1):85–100.
- Jamieson D (2015) The life and death of an amazon warehouse temp. *Huffington Post* .
- Johnson DS, McGeoch LA (1997) *Local Search in Combinatorial Optimization*, chapter The Traveling Salesman Problem: A Case Study in Local Optimization, 215–310 (London: John Wiley & Sons).
- Johnson ME, Brandeau ML (1996) Stochastic modeling for automated material handling system design and control. *Transportation science* 30(4):330–350.
- Kapner S (2015) How the web drags on some retailers: Costs of shipping, returns and handling pinch results. *The Wall Street Journal* URL <http://www.wsj.com/articles/how-the-web-drags-on-some-retailers-1417477790>.
- Kapner S (2016) Stores gear up for returns-and more shopping. *The Wall Street Journal* URL <https://www.wsj.com/articles/stores-gear-up-for-returnsand-more-shopping-1482769959>.
- KC DS, Staats BR (2012) Accumulating a portfolio of experience: The effect of focal and related experience on surgeon performance. *Manufacturing & Service Operations Management* 14(4):618–633.
- KC DS, Staats BR, Gino F (2013) Learning from my success and from others’ failure: Evidence from minimally invasive cardiac surgery. *Management Science* 59(11):2435–2449.

- Kliegl R, Smith J, Baltes PB (1986) Testing-the-limits, expertise, and memory in adulthood and old age. *Human memory and cognitive capabilities: Mechanisms and performances* 395–407.
- Krause EF (2012) *Taxicab geometry: An adventure in non-Euclidean geometry* (Courier Corporation).
- Kremer M, Debo L (2015) Inferring quality from wait time. *Forthcoming, Management Science* .
- Kremer M, Moritz B, Siemsen E (2011) Demand forecasting behavior: System neglect and change detection. *Management Science* 57(10):1827–1843.
- Kuhn HW (1955) The hungarian method for the assignment problem. *Naval research logistics quarterly* 2(1-2):83–97.
- Lapré MA, Mukherjee AS, Van Wassenhove LN (2000) Behind the learning curve: Linking learning activities to waste reduction. *Management Science* 46(5):597–611.
- Lapré MA, Tsikriktsis N (2006) Organizational learning curves for customer dissatisfaction: Heterogeneity across airlines. *Management science* 52(3):352–366.
- Malkevitch J (n.d.) Taxi! online, URL <http://www.ams.org/samplings/feature-column/fcarc-taxi>.
- Manzini R, ed. (2012) *Warehousing in the Global Supply Chain* (Springer-Verlag London), 1st edition.
- McClelland M (2011) A visit to the warehouse of soul-crushing sadness. *Mother Jones* URL <http://www.motherjones.com/rights-stuff/2011/07/ohio-warehouse-temps-unemployment>.
- McClelland M (2012) I was a warehouse wage slave: My brief, back-breaking, rage-inducing, dildo-packing time inside the online-shipping machine. *Mother Jones* URL <http://www.motherjones.com/politics/2012/02/mac-mcclelland-free-online-shipping-warehouses-labor>.
- Min H (2007) Examining sources of warehouse employee turnover. *International Journal of Physical Distribution & Logistics Management* 37(5):375–388.
- Mukherjee AS, Lapré MA, Van Wassenhove LN (1998) Knowledge driven quality improvement. *Management Science* 44(11-part-2):S35–S49.
- Noon CE, Bean JC (1993) An efficient transformation of the generalized traveling salesman problem. *INFOR: Information Systems and Operational Research* 31(1):39–44.
- Nothdurft HC (2002) Attention shifts to salient targets. *Vision research* 42(10):1287–1306.
- Petersen CG, Aase G (2004) A comparison of picking, storage, and routing policies in manual order picking. *International Journal of Production Economics* 92(1):11–19.
- Pisano GP, Bohmer RM, Edmondson AC (2001) Organizational differences in rates of learning: Evidence from the adoption of minimally invasive cardiac surgery. *Management Science* 47(6):752–768.
- Ramdas K, Randall T (2008) Does component sharing help or hurt reliability? an empirical study in the automotive industry. *Management Science* 54(5):922–938.
- Reichardt T (2011) Amazon: Leading the way through chaos. blog, URL <http://www.ssi-schaefer.de/blog/en/order-picking/chaotic-storage-amazon/>.

- Rose J, Gamble JG, eds. (2006) *Human walking* (Lippincott Williams & Wilkins Philadelphia), 3rd edition.
- Rosenholtz R, Li Y, Mansfield J, Jin Z (2005) Feature congestion: a measure of display clutter. *Proceedings of the SIGCHI conference on Human factors in computing systems*, 761–770 (ACM).
- Rosenholtz R, Li Y, Nakano L (2007) Measuring visual clutter. *Journal of vision* 7(2):17–17.
- Satariano A, Soper S (2015) What amazon wants for christmas: 100,000 temp. *BloombergBusinessweek* URL <http://www.bloomberg.com/news/articles/2015-12-10/what-amazon-wants-for-christmas-100-000-temps>.
- Song H, Tucker AL, Murrell KL, Vinson DR (2015) Public relative performance feedback in complex service systems: Improving productivity through the adoption of best practices. *Harvard Business School Research Paper Series* 16–043.
- Soshkin M (2015) Order fulfillment services in the us. IBISWorld Industry Report OD5504, IBISWorld, URL <http://clients1.ibisworld.com/reports/us/industry/ataglance.aspx?entid=5504>.
- Srivastava S, Kumar S, Garg R, Sen P (1969) Generalized traveling salesman problem through n sets of nodes. *CORS journal* 7:97–101.
- Staats BR, Gino F (2012) Specialization and variety in repetitive tasks: Evidence from a japanese bank. *Management Science* 58(6):1141–1159.
- Statista (2015) Holiday season sales as percentage of total retail e-commerce sales in the united states from 2012 to 2015. Online, URL <http://www.statista.com/statistics/283875/holiday-e-commerce-share-of-total-us-retail-e-commerce-sales/>.
- Tong J, Feiler D (2016) A behavioral model of forecasting Naive statistics on mental samples. *Management Science* Forthcoming.
- Tucker AL, Nembhard IM, Edmondson AC (2007) Implementing new practices: An empirical study of organizational learning in hospital intensive care units. *Management Science* 53(6):894–907.
- Turner S (2009) *Joint Multi-item Storage Location Assignment Capacitated Lot Sizing Problem*". Master's thesis, University of Windsor, paper 161.
- US Census Bureau (2016) Quarterly retail e-commerce sales: 1st quarter 2016. online, URL https://www.census.gov/retail/mrts/www/data/pdf/ec_current.pdf.
- Winston WL, VenKataramanan M (2003) *Introduction to Mathematical Programming* (Brooks/Cole Cengage Learning), 4th edition.
- Yang J, Shi X, Marchese M, Liang Y (2008) An ant colony optimization method for generalized tsp problem. *Progress in Natural Science* 18(11):1417–1422.