ICIS 2022 Proceedings User Behavior, Engagement, and Consequences

Dec 12th, 12:00 AM

The Colors of Performance – Assessing the Impact of Color-Coding on Worker Behavior in Retail Order Picking

Dominic Loske
REWE Logistics, dominic.loske@wiwi.uni-goettingen.de

Jannes Heinrich Diedrich Menck
University of Goettingen, jannes.menck@uni-goettingen.de

Henrik Lechte
University of Goettingen, henrik.lechte@uni-goettingen.de

Tim-Benjamin Lembcke
University of Goettingen, tim-benjamin.lembcke@uni-goettingen.de

Tiziana Modica
Politecnico di Milano, tiziana.modica@polimi.it

See next page for additional authors

Presenter Information
Dominic Loske, Jannes Heinrich Diedrich Menck, Henrik Lechte, Tim-Benjamin Lembcke, Tiziana Modica, and Matthias Klumpp

The Colors of Performance – Assessing the Impact of Color-Coding on Worker Behavior in Retail Order Picking

Completed Research Paper

Dominic Loske
University of Goettingen
Platz der Goettinger Sieben 3,
37073 Goettingen
dominic.loske@wiwi.uni-goettingen.de

Jannes Heinrich Diedrich Menck
University of Goettingen
Humboldtallee 3, 37073 Goettingen
jannes.menck@uni-goettingen.de

Henrik Lechte
University of Goettingen
Humboldtallee 3, 37073 Goettingen
henrik.lechte@uni-goettingen.de

Tim-Benjamin Lembcke
University of Goettingen
Humboldtallee 3, 37073 Goettingen
lembcke1@uni-goettingen.de

Tiziana Modica
Politecnico di Milano
Via Lamburschini 4, 20156 Milano
tiziana.modica@polimi.it

Matthias Klumpp
University of Goettingen
Platz der Goettinger Sieben 3,
37073 Goettingen
matthias.klumpp@uni-goettingen.de

Abstract

Advanced technologies are introduced in warehouse operations, rendering the interplay between human worker behavior and information systems (IS) a critical issue. We investigate how IS supports manual order picking by studying how visual color-coding information on picking locations provided through personal digital assistants accelerates search and picking tasks. Considering real-world data on a storage system where 20 dissimilar items are stored together at one picking location, we apply a log-logistic accelerated failure time model with N=112,672 picks performed by N=190 workers and find that color-coding accelerates the picking process by up to 17.28%. To increase the internal validity of our field-based examination, we conduct one VR experiment (N=29 participants) providing evidence for an acceleration of 23.74%, and one online experiment (N=178 participants) indicating an acceleration of 24.29%. Based on an innovative method of triangulation, we demonstrate how IS can influence picker behavior and discuss how to better design IT artifacts.

Keywords: Color-coding, worker behavior, order picking

Introduction

Warehouse workers are currently witnessing the introduction of various technologies that are rapidly transforming their working environments (Fragapane et al. 2021). Although robots are implemented to automate routine tasks (Autor 2015; Frey and Osborne 2017), 80% of all orders processed by warehouses are still picked manually (Boysen et al. 2021), and according to the latest statistics, more than one million workers are employed in the US warehousing industry (US Department of Labor 2022). Thus, warehouse workers still play a vital role in order picking.

In order to increase the performance of the order picking system, information systems (IS) assisting workers and facilitating order picking tasks are increasingly applied in warehousing (Glock et al. 2021). Despite the support received by IS systems, human workers' behavior accounts for a large share of performance deviation from the expected outcome, as supported by numerous studies in the research field of behavioral operations management (BeOM), e.g., Donohue et al. (2020) and Sun et al. (2022). BeOM studies how individual or group level behavior (e.g., of managers, workers, or customers) impacts decision-making, using this understanding to improve supply chain operations (Katsikopoulos and Gigerenzer 2013). With the increasing adoption of IS and technologies supporting decision-making in supply chain operations, recent BeOM studies suggested further investigation on how these instruments can be structured and integrated with human behavior to improve people's decisions (Fahimnia et al. 2019). In the context of order picking, such contributions could be highly relevant for major e-commerce platforms (e.g., Amazon or Alibaba) and brick-and-mortar retailers (e.g., Walmart or REWE Group) as order picking is directly linked to customer satisfaction and, on average, represents 55% of the total warehouse operating expenses (de Koster et al. 2007; Richards 2021).

In this landscape, we investigate how a specific IS could facilitate order picking tasks performed by warehouse workers. We are especially concerned with visual search and how IS can accelerate finding a particular item. Bravo and Farid (2004) and Rosenholtz et al. (2007) find that search time increases when items are nearby, or of similar color. Therefore, we explore how digital nudging through IS can accelerate picking tasks from racks where dissimilar items are stored close together in a single location.

From a theoretical viewpoint, we draw on nudge theory. Nudge theory originates from behavioral economics and proposes heuristics and biases that alter human behavior in decision-making without limiting their choices, by creating an environment attracting attention (Lembcke et al. 2019; Thaler and Sunstein 2021; Weinmann et al. 2016). Drawing on the empirical results of Batt and Gallino (2019), which show the influence of the color of the items on order picking time, our IS supports warehouse workers' visual search by highlighting the items to be picked with different colors (hereinafter called color-coding). The IS includes an information screen and a user interface (Alon et al. 1995). Our hypothesis states that the use of this IS facilitates order picking tasks performed by warehouse workers, thus reducing the order picking time. A mixed-method approach supported the development of our findings. In a first step, our hypothesis was tested through field-based research developed in collaboration with REWE Group, a major European brick-and-mortar grocery retailer operating several warehouses for perishable and non-perishable items. Here, a survival analysis was proposed, considering order picking time as our dependent variable, which is in line with existing research (Batt and Gallino 2019; Sun et al. 2022). Our independent variable was modeled as a dichotomous variable operationalizing whether a pick is performed with or without being nudged by the color-coding of our IS. We further control for nine variables quantifying the order picking process: Weight per stock-keeping unit, volume per stock-keeping unit, pick level, travel distance, number of picks, picker per batch, primary packages in secondary packaging, the experience of the order picker, and the load unit utilized. The model was empirically tested with an archival dataset comprising N=112,672 picks performed by N=190 order pickers between May and December 2021.

To elude the impact of unobserved confounding variables, increase the replicability of our results, and mitigate possible concerns about the internal validity of our field-based data examination, we conduct a triangulation approach. This includes two lab experiments following our field-based examination, with one online experiment (N=178 participants) and one virtual reality (VR) experiment (N=29 participants). The online study is designed as a between-subject experiment simulating the picking tasks using pictures of real-world warehouse shelves. Online experiments are a common method to study real-world behavior (e.g., Knemeyer and Naylor (2011)) that have gained popularity during the COVID-19 pandemic (Peyton et al. 2021). With the wide-spread use of low-cost VR technologies, VR has also become a valid tool for conducting experimental research (Cipresso et al. 2018). Hummel and Maedche (2019) call for research using tools like VR for nudging-related research. In our case, VR offers the possibility to maintain a controlled setting while presenting a close-to-real-life scenario to study participants since the participants' perception of the experimental setting can be handled through the sense of presence created by immersion (Innocenti 2017). In contrast to the online experiment, participants can't see the objects all at once on a screen but as if they were standing in front of a real-world rack. The study design is, like the online experiment, a between-subject experiment.

In sum, our *contributions* are as follows. We contribute to BeOM literature on order picking providing empirical evidence on the positive impact of color-coding heuristics on order picking time. This evidence is incorporated into an IS to support individual decision-making. We provide empirical evidence that the positive nudging impact of color-coding heuristics has a relevant effect size on order picking time, contributing to shedding light on the application fields of nudge theory. Finally, we propose and apply a new triangulation approach combining empirical data analysis, online modelling, and VR lab experiments that increase the internal validity of empirical research models dedicated to the interplay of worker behavior and IS.

Research Background

Our study combines and integrates multiple research streams. First, we relate to BeOM literature studying human behavior in warehousing. As a field that investigates the effects human behavior has on operations management, BeOM is recognized as a promising and relevant emerging research subject (Bendoly et al. 2010). Recent research has called for further investigations on how to structure technologies supporting decision-making processes, such as IS or other decision support mechanisms, to allow people to make better decisions (Fahimnia et al. 2019). BeOM literature focusing on order picking is scant. Extant studies have examined the effect of human learning and work characteristics on the automation of order picking (Loske 2022), analyzed human deviation from algorithmic prescription, proposing a new algorithm that incorporates these deviations to reduce them and improve the performance of bin-packing (Sun et al. 2022), and studied the effect of workers' experience on order picking time, developing an experience-based routing heuristic able to improve the order picking productivity by up to 3% (Batt and Gallino 2019). By studying how color-coding influences workers' performance in order picking, we extend the discussion by providing a novel perspective on IS capable of taking advantage of user behavior.

Second, we relate to nudge theory. Nudges are defined as "any aspects of the choice architecture that alters people's behavior in a predictable way without forbidding any options or significantly changing their economic incentives" (Thaler and Sunstein 2008, p. 6). Nudge theory assumes that choice architecture can be used to influence people's behavior, and it has been experimented with by researchers in different contexts to improve decision-making (Hummel and Maedche 2019). As displayed by Hummel and Maedche (2019) the effect size of nudge, computed as the difference between the value of the dependent variable of the treatment group and the control group, is affected by the context of application and the tool used for nudging people. So far, the context of the warehouse has been disregarded as an application setting for nudge theory. As nudges relate to human behavior in decision-making, we argue that warehouse operations can benefit from the introduction of nudge, since worker behavior plays a role in determining the performance of a warehouse. As we aspire to apply nudge theory to this new domain, we discuss how we contextualize the nudge and which tool we select as appropriate for our aim. When performing picking activities, the worker must search through a set of items to find the correct one. The searching activity, which is subjected to individual behavior (Loske 2022), could be complex and time-consuming, especially for inexperienced workers or systems with a high density of items (Batt and Gallino 2019). Thus, particularly designed nudges could be applied to help the worker's search. Among the different tools available for nudging (Sunstein 2014), we argue that graphic warnings, such as large fonts, bold letters, or bright colors best fit order picking search, since they are mechanisms used to capture people's attention. Evidence exists that colors for highlighting information could improve information search (Christ 1975). Moreover, recent research showed that colors affect the time to localize objects, regardless of a screen-based or a real-world setting (Huang 2008; Nuthmann and Malcolm 2016). Nowadays, colors are an important part of interface designs and are used for nudging users toward desired behaviors and perceptions (Bermejo Fernandez et al. 2021; Schneider et al. 2018). We, therefore, assume that color-coding nudges order picking tasks performed by warehouse workers, thus reducing the order picking time. This color-coding scheme could be implemented both in real-world settings or through an IS.

Third, we provide a methodological contribution for multidisciplinary research interested in evaluating the impact of user behavior in IS interactions. While analyzing archival data through parametric and non-parametric statistical methodologies is the state of the art, e.g., quantifying individual picker skills through multi-level modeling (Matusiak et al. 2017), the combination of field-based research and lab experiments remains scarce. Our results indicate that combining both can compensate for a majority of the respective weaknesses enabling evidence from real-world data to be validated through an isolated stimulus tested in different digital or virtual environments.

Research Setting

In collaboration with REWE Group, a leading European brick-and-mortar grocery retailer, and its logistics subsidiary REWE Logistics, we conducted an empirical field-based research project on the impact of IS design to improve order picking systems. We closely worked together with operations managers from a warehouse for non-cooled perishable items between May and December 2021. The objective was to evaluate how technologies shape and influence user behavior. When we identified that color-coding is a critical factor impacting the interplay between worker behavior and IS, we isolated the stimulus and tested it through one online experiment and one VR experiment. The goal of this triangulation approach is to empirically validate how nudging warehouse employees through color-coding influences worker behavior in manual order picking.

Empirical Field-based Research and Data Collection

To start off, we empirically evaluated the impact of color-coding in an empirical field-based research project in cooperation with REWE Group. Orders are picked in a manual picker-to-parts order picking system where pickers travel on the ground level to the picking location to retrieve stock-keeping units (Boysen et al. 2021). Each order picker uses one industrial truck for traveling in the warehouse and receives information on the order picking task from a personal digital assistant wirelessly connected to the warehouse management system.

Our field-based research concentrated on a high-density flow rack storage system, where individuals face the difficulty of finding one pick location among 20 equally-sized pick locations on a shelf of 2.40 meters in length and 2.00 meters in height. In order to accelerate the individual search processes for these picking locations, the operations managers integrated a logic where rack levels are colored for 55% (279 of 507) of the high-density flow rack storage systems in the warehouse. Each storage system consists of four levels with equal colors: The ground floor level (white color), level 1 at 0.60m above ground floor level (yellow color), level 2 at 1.20m above ground floor level (blue color), and level 3 at 1.80m above ground floor level (green color). The respective color is also presented on the personal digital assistant interface in front of each warehouse worker (see Figure 1).

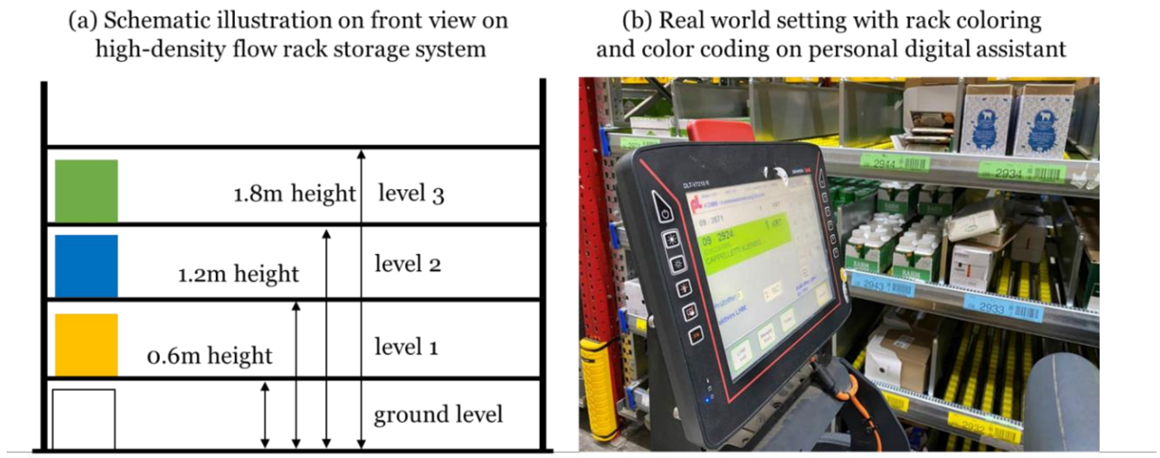

Figure 1. Research setting for field-based research

Most warehouse management systems (WMS) store extensive log data on order picking processes and capture them in a very detailed fashion. We utilize such empirical log data to construct a model capable of evaluating the accelerating and decelerating impact of color-coding on order picking performance. We implicitly include human factors through quantitative WMS data on the past performance of each order picker. Our dataset includes data on batch-ID, pick-ID, picker-ID, load unit-ID, article number, number of units picked, length, width, and height of secondary product packaging, the volume of secondary product packaging, the weight of product and secondary product packaging, timestamps of each pick, and slot address per pick.

Our initial dataset includes N=115,675 picks performed by 202 order pickers. Because we use real-world data, the logs are polluted for several reasons, e.g., personnel breaks or system breakdowns. Therefore, we exclude all picks lasting longer than 300 seconds as these have been identified as invalid data in the underlying scenario. Next, we control the speed of the industrial trucks and exclude all picks with a travel speed higher than 3.33 m/s. To control for the experience level of order pickers, we only integrate order pickers with at least 50 cumulative picks in the described setting. After cross-validating all data cleaning rules with the company, our final dataset comprises N=112,672 picks performed by 190 order pickers grasping items from 507 different pick locations in a high-density rack system.

Setup of Online Experiment and Data Collection

In order to avoid the impact of unobserved confounding variables in our field-based examination, we additionally set up an online experiment simulating the selection of products in a warehouse. Study participants were provided with 27 consecutive picking tasks for nine distinct shelves. The order was randomized for each participant. For each picking task, participants were provided a product number and had to select the corresponding product from a photograph of a real, almost-full warehouse shelf. The product selection was performed by clicking on either the product or the product's label. Each photograph comprised of 20 product slots (4 rows with 5 products each). The study participants were not provided with feedback regarding pick time or correctness but were asked to pick as correctly and quickly as possible. The experiment setup is depicted in Figure 2.

Figure 2. Research setting for the online experiment (translated from German) for the groups with no color (left) and matching color (right)

The online experiment was conducted as a between-subject design, with participants randomly assigned to three different groups. The first group was presented with non-colored shelf racks and labels, and the second and third groups with colored shelf racks and labels. For the second group, the product number's background in the description was colored the same color as the shelf rack holding the target product. For the third group, a color of a random, wrong shelf rack was used. Shelf photographs and pick tasks were identical for all groups except that the colored racks were altered to be black-and-white for the first group. Furthermore, artificial labels were added to the product slots to increase the visibility of the online experiment. The online experiment tracked pick time as the time between the task appearing on the screen and the participant selecting a product by clicking on the photograph of the warehouse shelf. In our data, we treat the first pick task as a test round and do not consider it in our analysis.

In addition to participating in the experiment, study participants had to self-rate their speed and correctness both through qualitative and quantitative questions. Participants could also indicate whether technical issues (e.g., photographs were not correctly displayed) occurred, and we discarded the data of 17 participants accordingly. Moreover, we excluded two color-blind participants from the data set. We also considered the number of successful picks and eliminated participants with fewer than 20 correct picks from the data set as we noticed that the number of successful picks far exceeds 20 for participants who genuinely attempted to complete the task. Furthermore, we removed all individual outliers with pick-times of over 20 seconds, arguing that those likely resulted from participants leaving the experiment temporarily, e.g., by switching to a different browser tab. These data cleaning steps resulted in a final number of 178 participants. The experiment participants were not recruited from the corporate field study, but included university students and participants from the international survey platform Prolific, specifically and additionally to extend our sample towards older participants (Peer et al. 2017).

Setup of Virtual Reality Experiment and Data Collection

We further conducted a VR experiment to create a close-to-real-life but still controlled setting. The simulation was created with Unity3D, a state-of-the-art game engine that enables developers to create realistic physical simulations (Unity 2022). The simulation was built for the Meta Quest 2, a standalone (no external computing power needed) VR device (Meta 2022). The design of the simulation closely mimicked an actual warehouse: the rack had the same heights, and the tasks were presented on a display on the left of the participants. The product objects to be picked were kept as generic blocks to reduce external effects.

The simulation allowed participants to move freely in a virtual room. The movement could be done physically to reduce possible motion sickness (Keshavarz and Golding 2022). Furthermore, participants were given a controller for each hand, which allowed them to grab objects from a rack by performing a grabbing motion and pressing the "grab" button. Additionally, the controllers were translated into virtual hands in the simulation, which also performed a grabbing motion, thereby creating presence, and leading to higher grades of immersion (Bowman and McMahan 2007).

Following a detailed explanation of the controls and the possibilities in the environment, the participants were able to complete a tutorial in which they had to take an object from a table and place it at a designated area. Following the tutorial, participants could start with the experiment, where they had to complete 20 picking tasks (with one product object to be picked each) in random order. For each task, participants had to grab the object and place it on a predefined location. We stored both the time until grabbing and the time until completing a pick by natively measuring them within the simulation. We further saved data on which object participants had to pick and whether the correct object was picked. Participants didn't get feedback on the time or correctness of their pick but were asked to pick as quickly and correctly as possible. The setup of the simulation is shown in Figure 3.

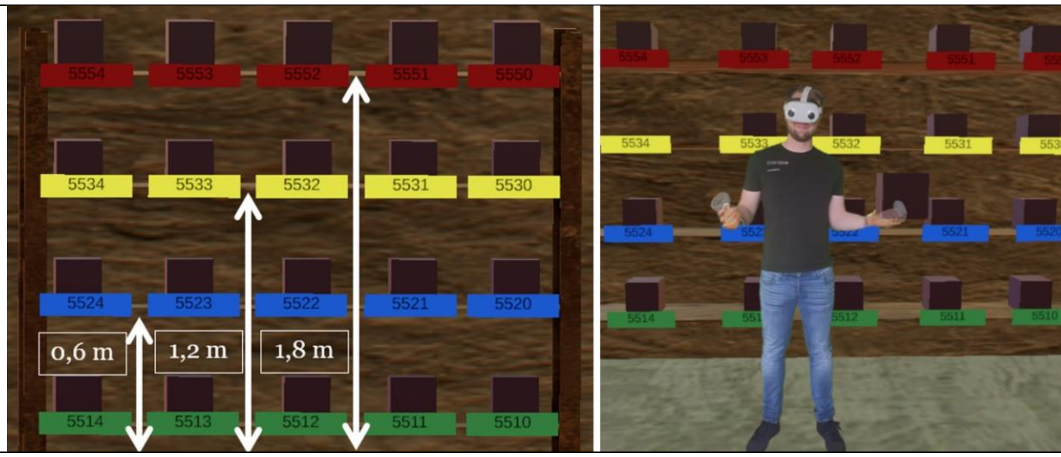

Figure 3. Research Setting VR Experiment

The VR experiment was conducted as a 2x1 between-subject experiment, with the participants randomly assigned to a group (color-coding/no color-coding). The experiment participants were not recruited from the field study, but separately. After the experiment, participants were asked to complete an identical questionnaire which was used for the online experiment. Because most of our participants had no VR experience, we excluded the first pick from each participant to minimize learning effects originating from the VR methodology.

Model Development for Empirical Analysis

To analyze all three triangulation data sources, we propose an event history analysis, also known as time-to-event analysis, or survival analysis, summarizing statistical models concerned with the probability and the duration until a given event occurs (Mills 2011). An event is formally defined as the instantaneous transition from the origin state to the destination state (Oud 2014), reflecting a broad conceptualization transferable to a large scope of scenarios in BeOM. In event history analysis, one essential methodological differentiation relates to the assumption regarding the effect of covariates. While proportional hazard models assume that covariates have a constant impact on the hazard function, accelerated failure time models assume an accelerating or decelerating impact (Greene 2018). In a nutshell, accelerated failure time models are regression models with different likelihood estimators than ordinary least-square regressions and use event time or survival time as the dependent variable (Mills 2011).

We transfer this logic to the human-computer interaction context and propose an accelerated failure time model to estimate the impact of coloring in manual order picking. This is inspired by the landmark paper of Batt and Gallino (2019). In our model, let T represent the time-to-event or survival time which we translate to the order picking context as picking time. T represents a random variable equal to or greater than zero ($T \geq 0$). In parametric survival models, T follows a particular distribution, e.g., exponential, Weibull, logistic, lognormal, or log-logistic. The choice for the parametric distribution assumed in the accelerated failure time model is made by comparing the model fit for various distributions e.g., the Akaike information criterion (AIC), Bayesian information criterion (BIC), or the Log-likelihood ratio (LL).

In the proposed econometric model, the time per pick is denoted as T, defined as the elapsed time between the beginning and end of a work process performed by one operator. Because accelerated failure time models are log-linear regression models for T, the basic model is a linear function of the covariate(s) in the form of $Y = \log(T)$ (Mills 2011). We define n independent predictor variables x_n and their corresponding regression coefficients β_n. Additionally, ε represents the error term assumed to have a particular parametric distribution. Our final model is as follows:

$$\begin{aligned}\ln(time\ per\ pick) = \quad & \beta_0 + \beta_1\, weight\ per\ SKU + \beta_2\, volume\ per\ SKU + \beta_3\, pick\ lelvel + \\ & \beta_4\, travel\ distance + \beta_5\, number\ of\ picks + \beta_6\, picks\ per\ batch + \\ & \beta_7\, number\ of\ packages\ in\ SKU + \beta_8\, experience\ of\ order\ picker + \\ & \beta_9\, picker + \beta_{10}\, dummy.roll\ cage + \beta_{11}\, dummy.full\ palette + \\ & \beta_{12}\, dummy.half\ palette + \beta_{13}\, dummy.color + \mathcal{E} \qquad (1)\end{aligned}$$

where we dummy-code as follows:

$$dummy.roll\ cage = \begin{cases} 1 & if\ the\ picker\ uses\ a\ roll\ cage, \\ 0 & otherwise. \end{cases} \qquad (2)$$

$$dummy.full\ palette = \begin{cases} 1 & if\ the\ picker\ uses\ a\ palette, \\ 0 & otherwise. \end{cases} \qquad (3)$$

$$dummy.half\ palette = \begin{cases} 1 & if\ the\ picker\ uses\ a\ half\ palette, \\ 0 & otherwise. \end{cases} \qquad (4)$$

$$dummy.color = \begin{cases} 1 & if\ the\ pickplace\ and\ screen\ utilizes\ color\ coding, \\ 0 & otherwise. \end{cases} \qquad (5)$$

Coefficients in our parametric accelerated failure time model are interpreted as follows: A positive coefficient indicates that the log duration time increases, leading to extended pick duration times. A negative coefficient indicates that the log duration time decreases, leading to shorter picking times. The regression coefficients β_n are parametrized by the following transformation (Mills 2011):

$$100\ (exp(\beta_n) - 1) \qquad (6)$$

Measures

Time per pick (DV): The clock starts when the order picker confirms to start a pick by pushing "next" on a personal digital assistant mounted on the industrial truck. The clock ends when the picker travels to the pick location, picks the stock-keeping units, and confirms the pick by pushing a symbol representing one of the load units on the personal digital assistant. Both timestamps are used to set the border of the total event time, which we utilize as our dependent variable. Time per pick is operationalized as a continuous variable.

Screen and Rack Coloring (IV): The independent variable of interest is the application of screen and rack coloring, allowing the interaction of the worker and IS through a personal digital assistant. 200 of the 507 places were colorized during the entire examination period, while 307 were not. Screen and rack coloring are operationalized as a dichotomous variable (0 = no coloring applied; 1 = coloring applied). We summarize the descriptive statistics of all variables in Table 1.

No.	Variable	Operationalization	Mean	SD.
DV	Time per pick	Continuous	23.24	25.09
IV	Screen and rack coloring	Binary dummy	0 = no coloring (43.34%), 1 = coloring (56.66%)	
CV1	Weight per stock-keeping unit	Continuous	3.07	2.67
CV2	Volume per stock-keeping unit	Continuous	7.16	5.13
CV3	Pick level	Continuous	1.749	0.61
CV4	Travel distance	Continuous	16.25	25.06
CV5	Number of picks	Continuous	1.04	0.34
CV6	Number of picks per batch	Continuous	76.04	51.19
CV7	Primary packages	Continuous	13.11	7.80
CV8	Experience of order picker	Continuous	527,571	56,525
CV9	Load unit	Ordinal, 3 stages	1 = roll cage (48.68%), 2 = palette (49.63%), 3 = half palette (1.69%)	

Table 1. Descriptive statistics for vocal variables

As control variables (CV) that serve to operationalize the order picking task, we utilize:

- (CV1) *Weight per stock-keeping unit:* The weight of stock-keeping units directly impacts human energy expenditure in manual order picking systems (Battini et al. 2016). We integrate the weight per stock-keeping unit in kilograms as a continuous variable to control for physical effort in manual order picking.
- (CV2) *Volume per stock-keeping unit:* Similar to the weight of stock-keeping units, article dimensions are relevant for the stacking process in manual order picking systems. We integrate the volume per stock-keeping unit in liters as a continuous variable to control the article dimensions impacting the complexity of the packing problem.
- (CV3) *Pick level:* We differentiate between four levels in the high-density flow rack storage system (ground floor level, level 1 at 0.60m from the floor, level 2 at 1.20m from the floor, and level 3 at 1.80m from the floor).
- (CV4) *Travel distance:* Picker traveling is one of the most time-consuming processes in manual picker-to-parts order picking systems (de Koster et al. 2007). Thus picker routing became an integral aspect of warehouse design issues (Masae et al. 2020). We integrate the travel distance per pick in meters through a distance matrix as a continuous variable to control for picker traveling time.
- (CV5) *Number of picks:* One batch includes picks from several picking locations. Multiple picks per picking location may be required, which we quantify as the number of picks and a continuous variable.
- (CV6) *Cumulative number of picks per batch:* After grasping stock-keeping units from the picking location, they are stacked on a roll cage. Herein, the filling level of the roll cage is relevant for the picking time. The more stock-keeping units on a roll cage, the more complicated the packing problem (Dowsland and Dowsland 1992; Dyckhoff 1990). We integrate the cumulative number of pickers per batch as a continuous variable to control the packing problem's complexity.
- (CV7) *Primary packages in secondary packaging:* One secondary package contains a certain number of primary packages to create a stock-keeping unit. We integrate the number of primary packages in one secondary package as a continuous variable to integrate the design of the product packaging system.
- (CV8) *Experience of order picker:* Research on learning effects in manual picker-to-parts order picking systems has proven that performance increases through experience. Therefore, we integrate the cumulative number of prior picks per picker-ID for each pick to control for the order picker's inceasing experience. We exclude all picker-IDs with less than 50 cumulative picks in the dataset. The experience of each order picker is operationalized as a continuous variable.
- (CV9) *Load unit:* Different load units are used within the setting including roll cages, regular palettes, and small palettes. We dummy code each load unit type as one dichotomous variable (load unit_roll cage: 0 = no roll cage; 1 = roll cage; load unit_palette: 0 = no palette; 1 = palette, load unit_half palette: 0 = no half palette; 1 = half palette).

To check for possible correlation effects of our dependent, independent, and control variables, we conducted a cross-correlation analysis and summarized the results in Table 2, indicating that it is valid to include all variables in our analysis.

ID.	DV	IV	CV1	CV2	CV3	CV4	CV5	CV6	CV7	CV8
DV	1.00***	0.06***	- 0.01***	- 0.14***	0.19***	0.05***	0.04***	- 0.16***	0.28***	- 0.15***
IV	0.06***	1.00***	0.07***	- 0.05***	0.21***	0.18***	0.01***	- 0.02***	0.11***	- 0.19***
CV1	- 0.01***	0.07***	1.00***	- 0.03**	- 0.02***	0.01***	- 0.02***	0.02***	- 0.00***	0.03***
CV2	- 0.14***	- 0.05***	- 0.03***	1.00***	- 0.16***	- 0.22***	- 0.08***	- 0.07***	- 0.18***	0.10***
CV3	0.19***	0.21***	- 0.02***	- 0.16***	1.00***	0.60***	0.15***	0.01***	0.30***	- 0.50***
CV4	0.05***	0.18***	0.01***	- 0.22***	0.60***	1.00***	0.01***	- 0.03***	0.24***	- 0.26***
CV5	0.04***	0.01***	- 0.02***	- 0.08***	0.15***	0.01***	1.00***	- 0.10***	0.16***	- 0.10***
CV6	- 0.16***	- 0.02***	0.02***	- 0.07***	0.01***	- 0.03***	- 0.10***	1.00***	- 0.13***	0.02***
CV7	0.28***	0.11***	- 0.00***	- 0.18***	0.30***	0.24***	0.16***	- 0.13***	1.00***	- 0.27***
CV8	- 0.15***	- 0.19***	0.03***	0.10***	- 0.50***	- 0.26***	- 0.10***	0.02***	- 0.27***	1.00***
CV9	- 0.21***	- 0.19***	- 0.01***	0.15***	- 0.37***	- 0.35***	- 0.05***	0.19***	- 0.29***	0.40***
Note: *p < 0.05; **p < 0.01; ***p < 0.001.										

Table 2. Correlation analysis for variables

Data Analyses and Results

Real-world Field-based Examination: Descriptive Analysis and Model Estimation

We first create one log-logistic accelerated failure time model with all control variables (Model 1) and one integrating the IS and rack coloring as the independent variable of interest (Model 2). Then, we utilize the identical model but integrate picker fixed effects through a mixed model where we allow one regression line for each of the 190 order pickers in the sample in Model (3) and Model (4). When using the AIC, BIC, and LL ratios of Model (2) and Model (4), we find that the model fit improves when we formulate a mixed-effects model (lower AIC and BIC, higher LL). Additionally, when we compare the AIC and BIC ratios of Model (3) and Model (4), we also find that integration of IS and rack coloring as our independent variable of interest improves the model fit (lower AIC and BIC, LL is not reliable here as it always increases when variables are added). When we transform β_1 in Model (4) through Formula (6), we note that compared to picks without Screen and rack coloring, color-coding can accelerate the picking process by up to 17.28% ($p < 0.001$). Table 3 summarizes all relevant results.

Dependent variable: Time per pick				
	Model (1)	**Model (2)**	**Model (3)**	**Model (4)**
IV: Color-coding		-0.1748*** (0.0057)		-0.1898*** (0.0058)
Weight per stock-keeping unit	0.0202*** (0.0007)	0.0115*** (0.0007)	0.0217*** (0.0007)	0.0123*** (0.0007)
Volume per stock-keeping unit	0.0307*** (0.0011)	0.0363*** (0.0011)	0.0307*** (0.0011)	0.0367*** (0.0011)
Level	-0.0410*** (0.0016)	-0.0399*** (0.0016)	-0.0396*** (0.0016)	-0.0395*** (0.0016)
Distance	0.0086*** (0.0001)	0.0082*** (0.0001)	0.0087*** (0.0001)	0.0083*** (0.0001)
Number of picks	0.1767*** (0.0077)	0.1823*** (0.0077)	0.1789*** (0.0074)	0.1838*** (0.0074)
Primary packages	-0.0028***(0.0003)	-0.0027*** (0.0003)	-0.0032*** (0.0003)	-0.0030*** (0.0003)
Number of picks per batch	0.0006*** (0.00005)	0.0006*** (0.00005)	0.0006*** (0.00005)	0.0007*** (0.00005)
Load unit, full palette	0.1045*** (0.0051)	0.0822*** (0.0051)	0.1148*** (0.0051)	0.0908*** (0.0051)
Load unit, half palette	0.1399*** (0.0173)	0.1775*** (0.0173)	0.1449*** (0.0164)	0.1854*** (0.0166)
Experience of order picker	-0.0003*** (0.000004)	-0.0003*** (0.000005)	-0.0003*** (0.000004)	-0.0003*** (0.000004)
Picker fixed effect	No	No	Yes	Yes
Constant	2.5161*** (0.0138)	2.6205*** (0.0141)	2.4714*** (0.0136)	2.5921*** (0.0140)
Number of pickers	190	190	190	190
Observations	112,672	112,672	112,671	112,671
AIC	889,214	888,278	886,305	885,232
BIC	889,329	888,403	888,241	887,177
Log-likelihood	-444,595	-444,126	-442,951	-442,414
Degrees of freedom	10	11	200	201
chi^2	28,618 ***	29,556 ***	27,850 ***	28,925 ***
Note: Robust standard errors in parentheses; *p < 0.05; **p < 0.01; ***p < 0.001.				

Table 3. Empirical results for log-logistic accelerated failure time model

To test the validity of our model, we check for several assumptions regarding the distribution of our dependent variable, including Weibull, Gaussian, logistic, and lognormal compared to the log-logistic form we applied. Since we find the lowest AIC and BIC values and the highest LL ratios for the log-logistic, the underlying model assumptions are proven to be valid.

Dependent variable: Time per pick					
	Log-logistic	**Weibull**	**Gaussian**	**Logistic**	**Log-normal**
IV: Color-coding	-0.1898*** (0.0058)	-0.1441*** (0.0065)	-3.7379*** (0.1703)	-3.5753*** (0.1174)	-0.1653*** (0.0063)
Control variables	Included	Included	Included	Included	Included
Picker fixed effect	Yes	Yes	Yes	Yes	Yes
Constant	2.5921*** (0.0140)	2.9731*** (0.0153)	15.6164*** (0.3903)	13.0843*** (0.2736)	2.5709*** (0.0148)
Number of pickers	190	190	190	190	190
Observations	112,672	112,672	112,671	112,671	112,671
AIC	885,232	907,572	1,016,383	956,952	894,205
BIC	887,177	909,518	1,018,328	958,898	896,151
Log-likelihood	-442,414	-453,584	-507,989	-478,274	-446,900
Degrees of freedom	201	201	201	201	201
chi^2	28,925.9100***	21,609.2500***	17,907.6000***	27,016.9600***	23,647.2800***
Note: Robust standard errors in parentheses; *p < 0.05; **p < 0.01; ***p < 0.001.					

Table 4. Check for statistical validity of distribution assumption for dependent variable

Robustness Checks: Random Sampling Split and Picker Section

To mitigate possible concerns about the validity of our models, we add two robustness checks commonly applied in empirical analyses of archival data in operations management (Yao et al. 2021) - a 50/50 random sample split and a random picker selection. Our results in Table 5 indicate that the direction and effect size of our independent variable of interest is neither dependent on specific observation points (stable β_1 for 50/50 sample split) nor biased by specific pickers (stable β_1 for random picker selection). The robustness checks also validate that the model fit improves for mixed-effects models integrating human workers together with color-coding into the econometric model.

Dependent variable: Time per pick				
	50/50 sample split	**50/50 sample split**	**random picker selection**	**random picker selection**
IV: Color-coding	-0.1699*** (0.0081)	-0.1853*** (0.0082)	-0.1651*** (0.0084)	-0.1624*** (0.0089)
Control variables	Included	Included	Included	Included
Picker fixed effect	No	Yes	No	Yes
Constant	2.6136*** (0.0198)	2.5823*** (0.0197)	2.3381*** (0.0204)	2.2812*** (0.0207)
Number of pickers	190	190	82	82
Observations	56,336	56,336	56,281	56,281
AIC	444,369	442,923	420,932	419,422
BIC	444,485	444,720	421,048	420,271
Log-likelihood	-222,171	-221,260	-210,453	-209,616
Degrees of freedom	13	203	13	95
chi^2	14,507 ***	14,247 ***	19,030 ***	18,945 ***
Note: Robust standard errors in parentheses; *p < 0.05; **p < 0.01; ***p < 0.001.				

Table 5. Robustness checks

The VR experiment: Descriptive Analysis and Model Estimation

Similar to the previous models, we apply a log-logistic accelerated failure time model to the data obtained through our VR experiments. We calculate the estimates for two dependent time variables, including (1) the time to touch the correct item (mean 2.90 seconds, standard deviation 2.45 seconds) and (2) the time to touch and pack the correct items (mean 7.56 seconds, standard deviation 6.71 seconds). The mathematical notation for the model with time to touch the correct item is:

$$\ln(time\ to\ touch) = \beta_0 + \beta_1\ experience\ of\ order\ picker + \beta_2\ picker + \beta_3\ task + \beta_4\ dummy.color + \varepsilon \quad (7)$$

where we dummy-code as follows,

$$dummy.color = \begin{cases} 1 & if\ the\ pickplace\ is\ color-coded, \\ 0 & otherwise. \end{cases} \quad (8)$$

The mathematical notation for the model with time to touch and pack the correct items is:

$$\ln(time\ to\ touch\ and\ pick) = \beta_0 + \beta_1\ experience + \beta_2\ picker + \beta_3\ task + \beta_4\ dummy.color + \varepsilon \quad (9)$$

where we dummy-code as follows,

$$dummy.color = \begin{cases} 1 & if\ the\ pickplace\ is\ color-coded, \\ 0 & otherwise. \end{cases} \quad (10)$$

Table 6 provides comparison of Model (6) and Model (7) as well as Model (9) and Model (10), where the model fit improves (lower AIC and BIC, higher LL) when we integrate picker fixed effects in a mixed-effects model. When we transform the estimates for color-coding, compared to picks without rack coloring, color-coding can (1) accelerate the time to touch the correct item by up to 21.53% ($p < 0.001$) (log-transformed β_1 in Model 7) and (2) accelerate the time to touch and pack the correct item by up to 23.74% ($p < 0.001$) (log-transformed β_1 in Model 10). Statistical results of the VR experiment are summarized in Table 6.

	Dependent variable: Time to touch the correct item			**Dependent variable:** Time to touch and pack the correct item		
	Model (5)	**Model (6)**	**Model (7)**	**Model (8)**	**Model (9)**	**Model (10)**
IV: Color-coding		-0.4775*** (0.0753)	-0.2424*** (0.0566)		-0.2828*** (0.0265)	-0.2710*** (0.0269)
Experience of order picker	-0.0220*** (0.0068)	-0.0211*** (0.0066)	-0.0209*** (0.0042)	-0.0252*** (0.0025)	-0.0246*** (0.0023)	-0.0214*** (0.0020
Picker fixed effect	No	No	Yes	No	No	Yes
Constant	1.0403*** (00826)	1.2932*** (0.0887)	1.4371*** (0.0593)	2.1455*** (0.0312)	2.2960*** (0.0319)	2.2485*** (0.0298)
No. of participants	29	29	29	29	29	29
Observations	578	578	578	578	578	578
AIC	2,409	2,389	2,145	2,668	2,584	2,520
BIC	2,431	2,415	2,293	2,690	2,668	2,610
Log-likelihood	1,209	-1,190	-1,039	-1,339	-1,286	-1,231
Degrees of freedom	3	4	33	3	4	33
chi²	10.2473***	49.2559***	45.0783***	89.8497***	194.9623***	184.3406***
Note: Robust standard errors in parentheses; *p < 0.05; **p < 0.01; ***p < 0.001.						

Table 6. Results of the VR experiment

The Online Experiment: Descriptive Analysis and Model Estimation

Finally, we present the results of our online experiment using fully computer-generated content. We set the time to select the correct items as our dependent variable (mean 3.907 seconds; standard deviation 2.099 seconds). We first test for the impact of creating a mixed-effects model allowing for one regression line per participant. The decreasing AIC and BIC, together with the increasing LL from Model (12) to Model (13) and Model (14), as shown in Table 7, prove that integrating workers in the econometric model is beneficial for the model fit. This is in line with our previous findings. We test the following model:

$$\ln(time\ to\ select) = \beta_0 + \beta_1\ experience\ of\ order\ picker + \beta_2\ picker + \beta_3\ correct\ color + \beta_4\ wrong.color + \varepsilon \quad (11)$$

where we dummy-code as follows,

$$correct\ color = \begin{cases} 1 & if\ the\ pickplace\ is\ colored\ in\ the\ correct\ color, \\ 0 & otherwise. \end{cases} \tag{12}$$

$$wrong\ color = \begin{cases} 1 & if\ the\ pickplace\ is\ colored\ in\ the\ wrong\ color, \\ 0 & otherwise. \end{cases} \tag{13}$$

Furthermore, we then test for the impact of two independent variables of interest: Model (13) investigates the impact of color-coding similar to the previous models examining real-world and virtual reality scenarios. Additionally, as we are interested in the impact of misleading information on work tasks, we estimate the impact of wrong color-coding as a second interdependent variable of interest in Model (14). When we transform the estimates, we find that color-coding accelerates the selection process by up to 24.29% ($p < 0.001$). In contrast, misleading information that we operationalize by coding with the wrong color decelerates the selection process by up to 17.37% ($p < 0.001$). Statistical results of the online experiment are summarized in Table 7.

Dependent variable: Time to select the correct item				
	Model (11)	**Model (12)**	**Model (13)**	**Model (14)**
IV1: Color-coding			-0.2782*** (0.0097)	
IV2: Coding in wrong color				0.1602*** (0.0118)
Experience of order picker	-0.0104*** (0.0007)	-0.0099*** (0.0006)	-0.0092*** (0.0006)	-0.0108*** (0.0003)
Picker fixed effect	No	Yes	Yes	Yes
Task fixed effect	No	Yes	Yes	Yes
Constant	1.3203*** (0.0120)	1.3153*** (0.0109)	1.4290*** (0.0110)	1.2845*** (0.0054)
No. of participants	178	178	178	178
Observations	4,613	4,613	4,613	4,613
AIC	17,739	16,883	16,458	16,769
BIC	17,771	18,041	17,623	17,933
Log-likelihood	-8,864	-8,262	-8,049	-8,204
Degrees of freedom	5	186	187	187
chi^2	206.2826**	228.2009***	755.2909***	329.3402**
Note: Robust standard errors in parentheses; *p < 0.05; **p < 0.01; ***p < 0.001.				

Table 7. Results of the online experiment

Discussion and Outlook

There is an emerging interest among operations management scholars in the behavior of human decision-making in socio-technical systems (Donohue et al. 2020). Although various normative models have been developed in the past decades of operations management research, essential decisions in operational business are made by humans, e.g., managers, workers, or customers. We contribute to the understanding of human actions and uncover the benefit of nudging human behavior through color-coding.

By exploring the impact that color-coding has on order picking time with a multi-method approach, we contribute to different literature streams. First, we extend BeOM literature on order picking, suggesting that integrating color-coding into an order picking system including racks and IS could support individual decision-making for searching activities, accelerating order picking time by up to 17.28%. Since order picking in retailing and e-commerce is mainly performed by seasonal workers with many part-time or inexperienced workers (Batt and Gallino 2019), having an effective system that improves the efficiency of less experienced workers could improve the overall order picking performance. Another remarkable finding is that wrong color-coding serving as a proxy variable for misleading work task description decelerates the selection process by up to 17.37% in the online experiment. Therefore, the impact of color-coding on the outcome of order picking systems is mainly dependent on the design of the IS interface and the rack system. Based on our findings regarding the interplay of worker behavior and IS, we hypothesize that IS design moderates the relationship between user behavior and the outcome of a socio-technical system. When the design is capable of nudging behavior, in our case, with the correct color-coding and the identified acceleration effects, the utilization of IS and the adoption of IT artifacts can lead to increased user performance and cost-saving behavior. On the contrary, poor design leading to the IS providing incomplete (no color) or misleading (wrong color) information may even have a large negative impact on the outcome of socio-technical systems, as shown in the order picking case. Our results prove that human-centered system design can improve the outcome of operations systems – a central hypothesis of Industry 5.0 (European commission 2021; Kadir and Broberg 2021; Vijayakumar et al. 2021).

Second, we contribute to nudge theory by extending it to the warehouse domain. This is something new in the literature. Our results prove that color-coding mechanisms integrated into the warehouse rack systems and the supporting IS can facilitate workers' searching activities in an effective way. In particular, the online and the VR Experiments proved that this type of nudge is particularly relevant in this context, with an average effect size (22.91% considering the searching activities for the online and the VR Experiments) higher than the median effect size (21%) of previous studies investigating different fields (Hummel and Maedche, 2019). These results further contribute to the discussion on the effectiveness of low-cost nudges.

Third, the study offers many methodological considerations. It provides interesting insights for researching real-life questions. In some cases, it might not be possible to collect real-world data without significant investments. With our threefold approach (field, online, and VR), we show that an online experiment might give the first indication for a possible effect, while a VR experiment might result in close-to-real-life insights. This paper also offers a strong method in terms of validity: while the field study has the strongest external validity, the VR Experiment has the strongest internal validity as it simulates the real picking task. Our results indicate that VR Experiments enable researchers to produce results with strong external validity while maintaining a strong internal validity. Because leading journals and scholars in operations management, e.g., Chandrasekaran et al. (2020), Chun et al. (2022), or Lam et al. (2022) propose to apply field-based research, field experiments, and intervention-based research for BeOM issues, our triangulation approach is meant to motivate researchers to combine lab experiments and field-based research, which are often applied in isolation. Our approach allows evaluating the effects of color-coding in three abstraction levels: No abstraction, VR (picking has to be done physically, the distance of the objects to pick is close to real-life), and online (picking non-physical objects on a screen). Finally, by showing how researchers can use VR as a tool for conducting research, we contribute to understanding and theorizing VR systems Wohlgenannt et al. (2020). As Hummel and Maedche (2019) suggested, we used a VR simulation to evaluate nudging aspects. The real-life data and the data obtained from the simulation yield similar results and support that VR experiments might be a valid way of researching nudging effects. We hope to inspire IS researchers to consider the use of VR simulations for their experiments as we argue that the results might be more transferable to the real world than the results of online experiments.

From a practical perspective, this study shows through a triangulation effort with three method applications that color-code nudging can have a positive impact on worker performances by reducing order picking time. This is something relevant since the order picking process is directly linked to customer satisfaction and accounts, on average, for 55% of the total warehouse operating expenses (de Koster et al. 2007; Richards 2021).

Conclusions and Further Research

In this paper, we present an evaluation of color-coding for visual search within warehouse order picking tasks. We show that incorporating a color-coding scheme connecting IS interfaces and the rack system can reduce the time for necessary search tasks in manual order picking, thereby reducing costs. By applying a log-logistic accelerated failure time model, we find an acceleration of 17.28%, which is grounded on real-life archival data for N=190 order pickers. Tests on the statistical significance, as well as the robustness of our results, confirmed our empirical findings. To exclude possible bias through non-observed effects and to increase the internal validity of our approach, we add two lab experiments. In a VR experiment, we find that color-coding can accelerate the selection and picking process of participants up to 23.74%, and in an additional online experiment, up to 24.29%. As the first empirical analysis of color-coding in manual picker-to-parts order picking systems, this study has limitations that offer several opportunities to investigate how to nudge user behavior, transfer our findings to customer scenarios, and investigate further learning effects.

First, our study was dedicated exclusively to color-coding and its impact on human order picking performance. However, visual search during order picking may be impacted by various other aspects, including for instance, the product color or the color of the background. Detecting pick locations or products in cluttered shelves is particularly difficult for workers because they have to check several potential locations

(Bravo and Farid 2004, 2008). Besides the aspect of order picking system design, the IS design providing information may also be relevant for order picking performance. Thus, further research might address the interplay of the identified color-coding effect with other human-computer interaction scenarios and digital interface communication issues such as font sizes, fonts types, dynamic representation (moving information), screen size and quality, and the combination of information provided by voice interactions with typical operational interfaces like scanning with personal digital assistants.

Second, our empirical results from field-based examinations and lab experiments prove that picking (real-world) and selection (virtual-world) tasks can be accelerated and are limited to order picking operations. Further research could center on customers as relevant actors in business contexts, not only in the context of BeOM but also for service operations management. Hence, our triangulation approach could be applied to an omnichannel retailing scenario where similar products like groceries or clothes, are purchased in a store (offline), through a webpage (online), or via a click-and-collect solution. A combination of field-based examinations (in a store) and lab experiments (simulating the homepage) could then compare the impact of promotion design or whether attention shifts to more salient targets (Nothdurft 2002; Rosenholtz et al. 2007).

Third, we observed a positive impact of experience as a key control variable in the field-based examination as well as in all lab experiments. The more cumulative experience users gain when interacting with the IS, the faster the picking and selection processes. When we log-transform the estimates from the field-based examination, we find that with each additional pick, the picking time decreases by 0.0299%. For the online experiment, the reduction is 1.07%. We find the steepest learning curve for the participants of the VR experiment where each additional pick decreases the time of the subsequent pick by up to 2.11%. Although the results are limited to order picking scenarios and, therefore, hard to generalize for operations management, learning effects through color-coding may vary depending on the IS. An examination of learning curves possibly depending on IS design could foster interesting insights on training issues for workers.

Finally, the existing setting with the applied method triangulation could be extended in the respective databases and sizes in order to establish a broader base for validation of the observed issues. This could provide a generally interesting cornerstone for IS research methodology as well as real-world IT artifact design for applications like those in the warehousing domain.

References

Alon, N., Yuster, R., and Zwick, U. 1995. “Color-coding,” *Journal of the ACM* (42:4), pp. 844-856.

Autor, D. H. 2015. “Why Are There Still So Many Jobs?: The History and Future of Workplace Automation,” *Journal of Economic Perspectives* (29:3), pp. 3-30.

Batt, R. J., and Gallino, S. 2019. “Finding a Needle in a Haystack: The Effects of Searching and Learning on Pick-Worker Performance,” *Management Science* (65:6), pp. 2624-2645.

Battini, D., Glock, C. H., Grosse, E. H., Persona, A., and Sgarbossa, F. 2016. “Human energy expenditure in order picking storage assignment: A bi-objective method,” *Computers & Industrial Engineering* (94), pp. 147-157.

Bendoly, E., Croson, R., Goncalves, P., and Schultz, K. L. 2010. “Bodies of knowledge for research in behavioral operations,” *Production and Operations Management* (19:4), pp. 434-452.

Bermejo Fernandez, C., Chatzopoulos, D., Papadopoulos, D., and Hui, P. 2021. “This Website Uses Nudging: MTurk Workers' Behaviour on Cookie Consent Notices,” *Proceedings of the ACM on Human-Computer Interaction* (5:CSCW2), pp. 1-22.

Bowman, D. A., and McMahan, R. P. 2007. “Virtual Reality: How Much Immersion Is Enough?” *Computer* (40:7), pp. 36-43.

Boysen, N., de Koster, R. B. M., and Füßler, D. 2021. “The forgotten sons: Warehousing systems for brick-and-mortar retail chains,” *European Journal of Operational Research* (288:2), pp. 361-381.

Bravo, M. J., and Farid, H. 2004. “Search for a category target in clutter,” *Perception* (33:6), pp. 643-652.

Bravo, M. J., and Farid, H. 2008. “A scale invariant measure of clutter,” *Journal of vision* (8:1), 23.1-9.

Chandrasekaran, A., Treville, S., and Browning, T. 2020. "Editorial: Intervention-based research (IBR)—What, where, and how to use it in operations management," *Journal of Operations Management* (66:4), pp. 370-378.
Christ, R. E. 1975. "Review and Analysis of Color Coding Research for Visual Displays," *Human Factors* (17:6), pp. 542-570.
Chun, Y., Harris, S. L., Chandrasekaran, A., and Hill, K. 2022. "Improving care transitions with standardized peer mentoring: Evidence from intervention based research using randomized control trial," *Journal of Operations Management* (68:2), pp. 185-214.
Cipresso, P., Giglioli, I. A. C., Raya, M. A., and Riva, G. 2018. "The Past, Present, and Future of Virtual and Augmented Reality Research: A Network and Cluster Analysis of the Literature," *Frontiers in Psychology* (9), p. 2086.
de Koster, R. B. M., Le-Duc, T., and Roodbergen, K. J. 2007. "Design and control of warehouse order picking: A literature review," *European Journal of Operational Research* (182:2), pp. 481-501.
Donohue, K., Özer, Ö., and Zheng, Y. 2020. "Behavioral Operations: Past, Present, and Future," *Manufacturing & Service Operations Management* (22:1), pp. 191-202.
Dowsland, K., and Dowsland, W. 1992. "Packing problems," *European Journal of Operational Research* (56:1), pp. 2-14.
Dyckhoff, H. 1990. "A typology of cutting and packing problems," *European Journal of Operational Research* (44:2), pp. 145-159.
European commission. 2021. "Industry 5.0: Towards more sustainable, resilient and human-centric industry," available at https://ec.europa.eu/info/news/industry-50-towards-more-sustainable-resilient-and-human-centric-industry-2021-jan-07_en.
Fahimnia, B., Pournader, M., Siemsen, E., Bendoly, E., and Wang, C. 2019. "Behavioral Operations and Supply Chain Management–A Review and Literature Mapping," *Decision Sciences* (50:6), pp. 1127-1183.
Fragapane, G., Koster, R. de, Sgarbossa, F., and Strandhagen, J. O. 2021. "Planning and control of autonomous mobile robots for intralogistics: Literature review and research agenda: Literature review and research agenda," *European Journal of Operational Research* (294:2), pp. 405-426.
Frey, C. B., and Osborne, M. A. 2017. "The future of employment: How susceptible are jobs to computerisation?" *Technological Forecasting and Social Change* (114), pp. 254-280.
Glock, C. H., Grosse, E. H., Neumann, W. P., and Feldman, A. 2021. "Assistive devices for manual materials handling in warehouses: a systematic literature review," *International Journal of Production Research* (59:11), pp. 3446-3469.
Greene, W. H. 2018. *Econometric analysis*, Harlow: Pearson Education Limited.
Huang, K.-C. 2008. "Effects of computer icons and figure/background area ratios and color combinations on visual search performance on an LCD monitor," *Displays* (29:3), pp. 237-242.
Hummel, D., and Maedche, A. 2019. "How effective is nudging?: A quantitative review on the effect sizes and limits of empirical nudging studies," *Journal of Behavioral and Experimental Economics* (80), pp. 47-58.
Innocenti, A. 2017. "Virtual reality experiments in economics," *Journal of Behavioral and Experimental Economics* (69), pp. 71-77.
Kadir, B. A., and Broberg, O. 2021. "Human-centered design of work systems in the transition to industry 4.0," *Applied ergonomics* (92), p. 103334.
Katsikopoulos, K. V., and Gigerenzer, G. 2013. "Behavioral Operations Management: A Blind Spot and a Research Program," *Journal of Supply Chain Management* (49:1), pp. 3-7.
Keshavarz, B., and Golding, J. F. 2022. "Motion sickness: Current concepts and management," *Current opinion in neurology* (35:1), pp. 107-112.
Knemeyer, A. M., and Naylor, R. W. 2011. "Using Behavioral Experiments to Expand Our Horizons and Deepen Our Understanding of Logistics and Supply Chain Decision Making," *Journal of Business Logistics* (32:4), pp. 296-302.
Lam, H. K. S., Ding, L., and Dong, Z. 2022. "The impact of foreign competition on domestic firms' product quality: Evidence from a quasi-natural experiment in the United States," *Journal of Operations Management*.
Lembcke, T.-B., Engelbrecht, N., Brendel, A. B., Herrenkind, B., and Kolbe, L. M. 2019. "Towards a Unified Understanding of Digital Nudging by Addressing its Analog Roots," *PACIS 2019 Proceedings*.
Loske, D. 2022. "Empirical evidence on human learning and work characteristics in the transition to automated order picking," *Journal of Business Logistics* (online first).

Masae, M., Glock, C. H., and Grosse, E. H. 2020. "Order picker routing in warehouses: A systematic literature review," *International Journal of Production Economics* (224), p. 107564.
Matusiak, M., de Koster, R. B. M., and Saarinen, J. 2017. "Utilizing individual picker skills to improve order batching in a warehouse," *European Journal of Operational Research* (263:3), pp. 888-899.
Meta. 2022. "Quest 2: Our Most Advanced New All-in-One VR Headset," *Meta Quest | Meta Store. https://store.facebook.com/quest/products/quest-2/.*
Mills, M. 2011. *Introducing survival and event history analysis*, Los Angeles: Sage.
Nothdurft, H.-C. 2002. "Attention shifts to salient targets," *Vision Research* (42:10), pp. 1287-1306.
Nuthmann, A., and Malcolm, G. L. 2016. "Eye guidance during real-world scene search: The role color plays in central and peripheral vision," *Journal of vision* (16:2), p. 3.
Oud, J. H. L. 2014. "Event History Analysis," in *Encyclopedia of quality of life and well-being research*, A. C. Michalos (ed.), Dordrecht: Springer, pp. 1-3 (doi: 10.1007/978-3-319-69909-7_953-2).
Peer, E., Brandimarte, L., Samat, S., and Acquisti, A. 2017. "Beyond the Turk: Alternative platforms for crowdsourcing behavioral research," *Journal of Experimental Social Psychology* (70), pp. 153-163.
Peyton, K., Huber, G. A., and Coppock, A. 2021. "The Generalizability of Online Experiments Conducted During the COVID-19 Pandemic," *Journal of Experimental Political Science*, pp. 1-16.
Richards, G. 2021. *Warehouse management*: *A complete guide to improving efficiency and minimizing costs in the modern warehouse*, London: KoganPage.
Rosenholtz, R., Li, Y., and Nakano, L. 2007. "Measuring visual clutter," *Journal of vision* (7:2), 17.1-22.
Schneider, C., Weinmann, M., and vom Brocke, J. 2018. "Digital nudging," *Communications of the ACM* (61:7), pp. 67-73.
Sun, J., Zhang, D. J., Hu, H., and van Mieghem, J. A. 2022. "Predicting Human Discretion to Adjust Algorithmic Prescription: A Large-Scale Field Experiment in Warehouse Operations," *Management Science* (68:2), pp. 846-865.
Sunstein, C. R. 2014. "Nudging: A Very Short Guide," *Journal of Consumer Policy* (37:4), pp. 583-588.
Thaler, R. H., and Sunstein, C. R. 2008. *Nudge: Improving decisions about health, wealth and happiness*, London: Yale University Press.
Thaler, R. H., and Sunstein, C. R. 2021. *Nudge*: *The Final Edition*, New York: Penguin Books an imprint of Penguin Random House LLC.
Unity. 2022. "Unity - Manual: Unity User Manual 2020.3," *https://docs.unity3d.com/Manual/index.html*.
US Department of Labor. 2022. "Industries at a Glance: Warehousing and Storage: NAICS 493," *available at https://www.bls.gov/iag/tgs/iag493.htm*.
Vijayakumar, V., Sgarbossa, F., Neumann, W. P., and Sobhani, A. 2021. "Framework for incorporating human factors into production and logistics systems," *International Journal of Production Research*, pp. 1-18.
Weinmann, M., Schneider, C., and vom Brocke, J. 2016. "Digital Nudging," *Business & Information Systems Engineering* (58:6), pp. 433-436.
Wohlgenannt, I., Simons, A., and Stieglitz, S. 2020. "Virtual Reality," *Business & Information Systems Engineering* (62:5), pp. 455-461.
Yao, Y., Duan, Y., and Huo, J. 2021. "On empirically estimating bullwhip effects: Measurement, aggregation, and impact," *Journal of Operations Management* (67:1), pp. 5-30.