Product Discovery and Consumer Search Routes: Evidence from a Mobile App [*]

Current version: July 2023

Luna Zhang[†] Raluca Ursu[‡] Elisabeth Honka[§] Oliver Yao[¶]

Abstract

When shopping online, consumers can reach a product detail page via multiple routes: by going through a category page (e.g., women's shoes), by directly typing the product name in the search field (e.g., Nike Women's Air Max), by going through a sales page (e.g., the shoes sale page), etc. Previous research has largely ignored how consumers choose between these routes and how, in turn, these routes affect which products consumers subsequently discover, search, and potentially purchase. Our novel panel data from a mobile shopping app contain detailed information on consumer browsing for sandals over a time period of six months. To capture consumers' decisions of not only what to search and buy, but also of the route through which to discover products, we build on the framework of Greminger (2021, 2022) and estimate a model of discovery, search, and purchase. We use our model to quantify preferences, discovery costs, and search costs, and show that product search costs are four times larger than product discovery costs. Via counterfactuals, we quantify the value of a search route to the online retailer and examine how app design changes influence product discovery and consumer purchase behavior.

Keywords: Consumer Search, Awareness, Product Discovery, Mobile Commerce, Apparel Industry.

JEL Classification: D83, L81, M31

---

[*]We would like to thank Giovanni Compiani, Rafael Greminger, Zhenling Jiang, Tai Lam, Ilya Morozov, and Stephan Seiler, seminar participants at Temple University, UC Berkeley, Frankfurt School of Finance and Management, Arizona State University, Florida International University, University of Texas at Dallas, University of Denver, Georgia Institute of Technology, American University, Bocconi University, UC Irvine, and participants of the 33rd POMS Conference, 21st International Industrial Organization Conference (IIOC), 2022 Marketing Science Conference, the 11th Workshop on Search and Switching Costs, the Tilburg X-mas Camp 2022, the 2021 Conference on Information Systems & Technology, and the 2021 Workshop on Information Systems and Economics for their comments. All errors are our own.

[†]University of Washington Tacoma, xyzhang5@uw.edu.

[‡]New York University, rmu208@stern.nyu.edu.

[§]University of California Los Angeles, elisabeth.honka@anderson.ucla.edu.

[¶]Lehigh University, yuy3@lehigh.edu.

1 Introduction

E-commerce represented about 14% of total U.S. retail sales in 2020 and is expected to grow to 24% by 2025.[1] Online retailers typically sell hundreds if not thousands of products in a category. For example, the apparel retailer Zara carries over 700 women's blouses and over 1,000 women's pants; Amazon lists over 700 trampolines and over 50,000 tea cups for sale; and Wayfair has nearly 19,000 dining tables and over 1,000 bathrobes in stock.[2] In many categories in which product design plays a central role and/or which are infrequently purchased, consumers are usually not aware which products are available for purchase and what each product's characteristics are prior to visiting an online retailer. As a result, they spend time discovering products before investigating a few in detail and potentially making a purchase. Hence, product discovery is an important component of the overall shopping process and assuming that consumers are aware of all available products does not appropriately reflect actual consumer behavior.

The retailer's website design plays an important role in the consumers' online shopping experience. Previous literature has shown that website design decisions affect consumers' shopping behavior on large-screen devices, such as desktops, laptops, and tablets. For example, Ngwe, Ferreira, and Teixeira (2019) showed that making it more cumbersome to find discounted products on a website can increase an online retailer's margins. Bairathi, Zhang, and Lambrecht (2022) show that adding a retailer badge, e.g., "Amazon's Choice," can increase sales of both products that received the badge and products that did not receive the badge. Website design decisions are likely even more consequential on small-screen devices, i.e., mobile phones. UI/UX design companies routinely recommend that mobile apps follow the 3-tap rule, i.e., consumers should be able to get to any product detail page in no more than three taps/clicks, and stress the importance of horizontal filtering.[3] Both

[1] https://www.smartinsights.com/digital-marketing-strategy/online-retail-sales-growth/.
[2] All websites were accessed in the U.S. on May 8th, 2023.
[3] See, e.g., https://zesium.com/top-10-tips-for-better-ecommerce-mobile-app-ux-design/,

recommendations exemplify the importance of website design for consumer shopping behavior as consumers can take different routes to reach a product list and product detail pages.

In this paper, we study an empirical setting in mobile commerce in which product discovery plays an important role. Specifically, we investigate how consumers discover, search, and purchase products on the mobile app of an apparel retailer and how the app design, more precisely the different ways of discovering products, affects consumer shopping behavior. Figure 1 illustrates the type of consumer behavior on the app we aim to describe. When a consumer first opens an app, she sees the homepage, which commonly shows a collection of products from multiple categories. Given an interest in a certain category, e.g., sandals, the consumer can reach a list of products in this category through different *search routes*. In our example, the consumer can choose between two search routes: the category (e.g., women's sandals) and the sales page (displaying promoted products). Choosing a search route, e.g., the category page, reveals a list of products, with four products typically displayed on one screen. At this point, the consumer has three options: (i) click on one of these four products, (ii) scroll down to discover four additional products on a new screen, or (iii) switch to using a different search route, i.e., the sales page. We define "discovery" as actions (ii) and (iii) above (scrolling down to see additional products or switching to a different search route to find additional products) and "search" as action (i): clicking on a discovered product to be directed to its product detail page. In our example, after scrolling down to see more products, the consumer decides to click on one of the products displayed on the category page and subsequently to buy.

===========================

Insert Figure 1 about here

===========================

The empirical search literature in marketing and economics has extensively studied consumers' online browsing of product detail pages and their subsequent purchase behavior

`https://www.softermii.com/blog/19-ux-design-tips-for-shopping-app-with-examples#form`, `https://stormotion.io/blog/top-6-ecommerce-mobile-app-design-tips-for-successful-sales/`.

either under the assumption that consumers are aware of all products available on a website or under the assumption that consumers are aware of all products on the currently viewed product list page (see, e.g., Koulayev 2014; Ursu 2018). Assuming that consumers are aware of all products an online retailer sells does not leave room for product discovery. Assuming that consumers are only aware of products on the currently viewed product list page is a weaker assumption. However, it does not answer the question of *how* consumers arrive at a product detail page, e.g., by going through a category, a recommendation or a sales page, and how this decision influences which products consumers discover, search, and potentially purchase.

We model how consumers discover, search, and purchase products using an adapted version of the theoretical framework developed in Greminger (2021) with an empirical application described in Greminger (2022) (the "Greminger framework" in the following). The Greminger (2021) model extends previous models based on Weitzman (1979) that assume that consumers are aware of all available products, but do not possess all relevant information (e.g., price, match value) about them and therefore need to search. In contrast, the Greminger framework allows consumers to discover products they are not aware of, while, at the same time, deciding which products to search and whether to buy. In other words, in this model, consumers are initially endowed with knowledge of a set of products and need to decide whether to discover more products, to search among already discovered products, or to stop and buy one of the searched products. Consumers can discover products through different discovery technologies, which correspond to the different search routes consumers can take to discover products in our empirical application. For this setting, Greminger (2021) derives optimal decision-making rules for all choices consumers make: which products to discover, which discovered products to search and in what order, what and whether to buy. These rules are characterized by three reservation values (i.e., index rules), which indicate when discover, search, or purchase decisions are optimal.

The model we estimate differs from the Greminger framework (more precisely its empirical application in Greminger 2022) in four main aspects. First, in our model, consumers can choose to discover products using multiple instead of only one search route and these search routes may contain the same, overlapping, or different sets of products. For example, while the category route typically contains all products in a category, the sales route only contains a subset of all products (those which are on sale). Second, because only four products are shown on each mobile app screen and because of a data limitation, in contrast to Greminger (2022), we do not take the ranking of products on a product list page into account. Third, our estimation approach is different. We use a smoothed frequency simulator and the lookup method to recover reservation utilities (Kim, Albuquerque, and Bronnenberg 2010), while Greminger (2022) derives a GHK simulator and estimates reservation utilities from pre-specified expressions. And lastly, our estimation approach allows us to directly recover all model primitives, i.e., preference parameters as well as discovery and search costs, whereas Greminger (2022) is not able to estimate discovery and search costs directly (which are not needed to answer his research question).

We use panel data from the mobile app of an apparel retailer to estimate our model. For a random sample of consumers, we observe every click they made in sessions in which they searched for women's sandals, our focal category, during a period of six months in 2017-2018. A nice feature of our data set lies in its panel nature: 48% of consumers are observed in more than one session. These 48% of consumers conduct, on average (median), 8 (5) sessions. In this mobile app, consumers can choose among six search routes (e.g., category page, sales page, search function). On average, consumers use 1.3 search routes, discover 21 products, and search 2 products in a session. Consumer behavior varies by search route. For example, consumers who use the category page or search function are most likely to make a purchase. Consumers who look at the category page or recommendations search more products than consumers who navigate the app using another search route.

Our estimation results show that consumers are sensitive to prices: they are more likely to search and purchase lower-priced sandals and sandals that are on promotion. These marketing mix decisions also affect other consumer decisions: lower prices or more promotions encourage consumers to search and purchase more, but lead to fewer discovered products. The intuition for this result is as follows: in the presence of a price reduction, consumers are more likely to find a product they want to (search and) purchase sooner and therefore terminate the process earlier than without the price reduction, which implies they discover fewer products overall. Further, we find discovery costs are substantially lower (approximately four times) than search costs. Therefore, consumers frequently search only a small subset of the discovered products and discover a relatively large number of products.

Via counterfactuals, we quantify the (marginal) value of a search route and evaluate the effects of changes in the design of the app. First, we measure the value of a search route (the sales page) by removing it from the app and predicting consumer behavior in this new environment. We find that the number of discoveries, searches, and purchases decreases, i.e., consumers do not "simply" switch to a different search route. Purchases decrease by 4% and revenue decreases by 4.2% when the sales page route is removed. The 4.2% of the revenue represents the (marginal) value of the search route to the retailer. And second, we examine the effects of changing the number of products displayed on a screen without needing to scroll down. By changing the number of displayed products, the app can affect the number of products consumers discover with every such decision. Currently, four products are displayed to consumers on a mobile screen. We predict consumer behavior using our model when this number is changed to two or six products per screen. We find that showing two products per screen decreases purchases by 8.4%, while showing six products per screen increases sales by only 0.2%. Since increasing the number of products displayed on a screen to six likely reduces the readability (which our model does not account for), while only marginally increasing sales, we interpret these results as evidence that displaying four products on a screen may be optimal for the retailer.

Our paper makes the following three contributions. First, our paper is among the first ones to model not only how consumers browse within a product list page (see, e.g., Chen and Yao 2017; Ursu 2018; Morozov et al. 2021), but also how they arrive at a product list page, i.e., through which search route. Modeling consumer choices of search routes allows us not only to quantify both discovery and search costs, but also to learn more about consumer preferences and how website designs affect consumer decisions. This is also important from a managerial perspective as it allows retailers to better understand what is important to consumers early in the shopping process.

Second, previous literature has explored the impact of website attributes, such as ease of use, navigability, and delays, on consumer shopping behavior (e.g., purchase and retention) primarily via surveys and lab experiments (Venkatesh and Agarwal 2006; Wells, Valacich, and Hess 2011). Research using observational data and research on the design of mobile apps are limited. Our study is among the first to explore the impact of mobile app screen design on consumer search and purchase decisions using secondary data.

And finally, despite the rapid growth of mobile commerce, little is known about consumer shopping behavior on mobile apps. Bang et al. (2013) identify two channel features – usability and ubiquity – that differentiate mobile and traditional online channels. Also, Ghose, Goldfarb, and Han (2013) suggest that search costs may be higher on mobile devices than on PCs. We contribute to the understanding of consumer behavior on mobile apps via descriptives as well as a quantification of discovery and search costs for this shopping channel.

The remainder of this paper is organized as follows. In the next two sections, we discuss the relevant literature and introduce our data. In Sections 4 and 5, we introduce our model, estimation approach, and identification strategy. We present and discuss our estimation and counterfactual results in Sections 6 and 7. Finally, we conclude in Section 8.

2 Relevant Literature

Our research is related to four streams of literature on product discovery and awareness, consumer search, mobile commerce, and app design. In the following, we review the relevant literature and delineate the positioning of our research vis-à-vis the findings from extant research.

It is well-known that consumers go through several stages (awareness, consideration, and purchase) in their shopping process before making a buying decision. As more granular data has become more available, researchers have begun to incorporate the consideration stage, often via a consumer search model, into demand models (e.g., Honka 2014; Ursu, Wang, and Chintagunta 2020) instead of maintaining the perfect information assumption that consumers are aware of and consider all available alternatives. A few papers have also included the awareness stage. For example, Honka, Hortaçsu, and Vitorino (2017) and Morozov (2023) model all three stages of the purchase process.[4] In these papers, awareness is viewed as a rather passive process in which consumers become aware of products due to advertising or personal characteristics. Greminger (2021) also builds a model in which all three stages of the purchase process are included. However, in contrast to Honka, Hortaçsu, and Vitorino (2017) and Morozov (2023), he models the set of products consumers discover (i.e., become aware of) as the outcome of an optimization problem. In other words, Greminger (2021) derives the optimal decision making strategy for a consumer who needs to determine which products to discover, to search, and to purchase at every step. In this paper, we use Greminger (2021)'s results on the optimality of consumer decisions in all three stages of the consumer purchase process to estimate consumer preferences as well as search and discovery costs. We extend Greminger (2021)'s model and apply it to a novel setting in which consumers discover products via multiple available search routes.

Sequential search models à la Weitzman (1979) have been widely used to study online consumer browsing behavior. For example, Kim, Albuquerque, and Bronnenberg (2010) and Kim, Albuquerque, and Bronnenberg (2017) study consumer shopping for camcorders on amazon.com, Koulayev (2014), Chen and Yao (2017), and Ursu (2018) investigate online hotel bookings, and Morozov (2023) examines consumer shopping for computer hard drives

[4] Morozov (2023) refers to the first stage in the consumer's purchase process as "category consideration." The author views awareness as part of category consideration.

using this model. More recently, researchers have worked on extending the Weitzman (1979) model to describe additional empirical patterns. For example, Gardete and Anthill (2020) allow the match value to be correlated with observable product characteristics, Ursu, Zhang, and Honka (2023) let consumers take breaks while searching, and Hodgson and Lewis (2021) allow consumers to update their beliefs based on already-searched products resulting in search path dependence. This paper falls into the group of papers extending the Weitzman (1979) model: we model product discovery via multiple routes in addition to consumer search.

The third stream of literature examines shopping and purchase behavior on mobile apps. Few papers have studied consumer browsing on mobile devices via the search framework. Ghose, Goldfarb, and Han (2013) examine consumer search for brand-related content on a microblogging website and suggest that search costs are higher on mobile devices than those on PCs. However, due to data limitations, the authors cannot estimate search costs and deduce them from posting recency. Zhang, Jiang, and Che (2019) develop a structural search model in which consumers first choose whether to search on a mobile device or a desktop and subsequently determine their search sets via a simultaneous search step. Using data from a field experiment on location-based mobile promotions, Fang et al. (2015) find these promotions to facilitate users' access to and retrieval of mobile promotion information, thus affecting consumer information search and planning of a future purchase. With data on consumer shopping in both the mobile channel and the traditional online channel, Wang, Malthouse, and Krishnamurthi (2015) find that mobile shoppers increase their order frequency and size due to the ubiquity of mobile devices. And lastly, Xu et al. (2017) exploit a natural experiment of the iPad app introduction in Alibaba's e-commerce market and note that the use of tablets increases consumer search for products and impulse purchase behavior. We add to this group of papers by estimating a sequential search model using data from a mobile app and by being one of the first papers that estimates search cost and the first to quantify discovery costs on mobile devices.

The fourth stream of literature examines human-computer interface design, such as the impact of website or mobile app attributes on consumer behavior. Venkatesh and Agarwal (2006) examine how website usability attracts and transforms visitors into customers using survey data. Through three lab experiments, Wells, Valacich, and Hess (2011) manipulate website quality and examine its impact on consumers' perceptions of product quality and online purchase intentions. Chen and Hitt (2002) collect data on online brokerage and find that system usage measures and system quality are associated with reduced consumer switching. Ngwe, Ferreira, and Teixeira (2019) conduct two field experiments on website design that vary the level of search frictions and investigate consumer online shopping behavior. Research on the impact of mobile app design choices is very limited. Using consumer browsing and purchase data from a mobile shopping app, Zhang, Cui, and Yao (2023) examine the impact of different information flows and functionality associated with mobile operating systems and app version designs on consumer impulse purchases. In this paper, we contribute to the understanding of the effects of app design changes on the entire purchase funnel through counterfactual analyses.

3 Data

3.1 Data Overview

Our data come from a mobile e-commerce retailer which sells consumer goods in 90 countries. The retailer was founded in 2012 and sells more than a million items in a variety of categories, including men's and women's apparel and footwear, beauty products, kids' and maternity products, electronics, home and living, sports, etc. via its mobile app. Most of the products are private label.

We focus on the nine countries with the most website activity (Saudi Arabia, United Arab Emirates, Indonesia, Kuwait, Jordan, Egypt, Oman, Qatar, and the Kingdom of Bahrain) and on women's sandals, the most frequently shopped category. We observe all clickstream activities on the retailer's mobile app for a random sample of consumers between September 2017 and February 2018 (six months). The clickstream activities include, among other things, visits of different page types, such as home page or category page, clicks on product detail pages, shopping cart additions, and time stamps for all activities. Further, the data encompass other consumer characteristics, such as consumers' IP addresses and app language. And lastly, the retailer also made product descriptions and regular prices available to us. We describe the data cleaning process in Web Appendix A, which resulted in a sample of 3,621 consumers who will be the focus of our study.

Given the clickstream nature of our data, one limitation is that we do not observe the products the consumer viewed but did not click on. In other words, we do not observe the products a consumer discovered on a product list page but did not search. We share this limitation with prior work using clickstream data (De los Santos, Hortaçsu, and Wildenbeest 2012; Bronnenberg, Kim, and Mela 2016; Chen and Yao 2017). To address this limitation, we reconstruct the data on the products the consumer viewed but did not click on.

Three features of the retailer's app aid us in the reconstruction process. First, the retailer does not customize which products are shown to a consumer and in which order, i.e., all individuals see the same products in the same order conditional on the same search route. Second, the set of available products does not change frequently. Products in the women's sandal category generally only change every season (i.e., every 3 - 4 months, rather than daily or weekly). And third, the retailer does not target prices or price promotions to individual consumers, i.e., all consumers see the same price for a specific pair of sandals. Furthermore, previous research has shown that the order in which products are shown to consumers has a large effect on consumer search behavior (Ursu 2018). In other words, products ranked first are more likely to be searched more frequently. We combine this insight from prior work and the three previously discussed app features to reconstruct the set and order of displayed products on product list pages using weekly search frequencies.[5]

[5]We also reconstructed the set and order of displayed products using monthly search frequencies and the reconstructed data exhibit similar characteristics. For example, the average differences in search order

We use all 3,621 consumers to calculate these weekly search frequencies for each search route separately. Then, we rank the products by their weekly search frequencies in a decreasing order, i.e., the most frequently searched product is ranked first, the second most frequently product is ranked second, etc., for each search route separately. Recall that a consumer is always shown four products on a screen. We assume that the first four most frequently searched products are shown on the first screen, the second four most frequently searched products are shown on the second screen, etc. Using the set of clicked products, we then determine how many screens a consumer viewed and therefore the set of discovered products. For example, if a consumer clicked on the fifth displayed product, then she would have discovered at least two sets of four products, i.e., the first eight products displayed on two screens. The details of the data reconstruction are available in Web Appendix A.

3.2 Data Description

Our data contain the browsing behavior of 3,621 consumers for women's sandals, the most frequently searched category, in 9,128 sessions. In total, 203 sessions resulted in a purchase. Furthermore, 48% of consumers visited the app in multiple sessions during the study period with an average and a median of 8 and 5 sessions, respectively, for consumers with at least two sessions.

Figure 2 shows the distributions of the number of search routes, the number of discovered products, and the number of searches in a session. All three distributions exhibit considerable variation. The average number of search routes a consumer uses during a session is 1.27 (median of 1) and the average number of searches a consumer makes in a session is 1.88 (median of 1). Based on the reconstructed data, the average (median) number of discovered products is 21.22 (16).

percentiles between the reconstructed data based on weekly and monthly frequencies are 1.4% and 2.1% for the category and sales routes, respectively.

==========================

Insert Figure 2 about here

==========================

In Table 1, we present descriptive statistics for several product features – once for all available products and once for all products consumers clicked on. The average regular price of a pair of sandals is 37 in local currency (around US $10) and 2% of the sandals are on promotion at any point in time. Not surprisingly, consumers tend to click on products that are cheaper and more likely to be on price promotion. We also display variables describing a consumer's search history before the consumer's current session. The variable "Searched Sandals Past Week" indicates whether a consumer searched a pair of sandals in the past week. The variable "Searched on Sales Page Past Week" captures whether a consumer used the sales route for any product (including sandals) in the past week. Although the statistics for these two variables look similar, only 20% of consumers are in both groups, i.e., searched sandals via the sales route in the week prior to the current session.

==========================

Insert Table 1 about here

==========================

Table 2 lists the six different search routes (and the home page) in our data and provides descriptive statistics for clicked and for purchased products conditional on having reached them via a specific route.[6] Consumer behavior varies depending on the search route they take. For example, consumers who go through the category or recommendations routes search more products on average. Consumers are most likely to make a purchase if they go through the category or search function routes. Consumers click on cheaper and more likely to be

[6] "Homepage" is the home screen that consumers see when they open the mobile app. "Category" is the category page. "Sales" is the sales page. "Search Function" describes the route used to look for a specific term by typing it into the search box and subsequently seeing a list page matching this search query. "Recommendations" relates to discovering products via recommendations available on the home page, on product pages, etc. "Product Page" describes the situation in which a consumer directly moves from a product page to another product page (not based on a recommendation). "Other" contains all other routes, such as discovering products through the consumer's account page, on the shopping cart page, etc.

price promoted products when they use the sale or other routes. Across all search routes, the average price of purchased products is lower than the average price of searched products.[7]

==========================

Insert Table 2 about here

==========================

4 Model

4.1 Utility, Discovery, and Search

Our goal is to model three decisions consumers make: (i) whether to discover additional products by choosing a (new) search route or by choosing to view more products in the current search route, (ii) whether to search an already-discovered product, and (iii) whether to purchase a searched product. To achieve this, we build on the Greminger framework which models discovery, search, and purchase decisions jointly. In our model, consumers can discover products through several different discovery technologies, which correspond to the different search routes consumers can take to discover products in our empirical application.

After opening the app and choosing an initial search route (e.g., the category page for women's sandals), the consumer is exposed to a set of products (e.g., four products) displayed as a product list on her screen. This set of products constitutes her initial awareness set, i.e., initial set of discovered products. The consumer then has to decide whether she wants to discover another set of products using any search route, whether she wants to search any of the products she has already discovered or whether she wants to terminate the process (by either purchasing a product among the searched ones or by taking the outside option of no purchase). Product discovery allows the consumer to expand her awareness set. The

[7]We note that the average prices of purchased products are calculated using a relatively small number of observations (203 purchases in total).

consumer can only search a product she has discovered and a larger awareness set allows the consumer to search among more products. Having searched more products allows the consumer to make her final choice among more alternatives.

Formally, consumer $i = 1, ..., N$ derives utility from buying product $j = 1, ..., J$ that equals

$$u_{ij} = \delta_{ij} + \epsilon_{ij}. \tag{1}$$

The utility of the consumer comes from two sources: δ_{ij}, her utility from product characteristics revealed through product discovery, and ϵ_{ij}, her utility uncovered through search. Before a discovery step, the consumer does not know (i) which products are going to be shown to her on a product list page and (ii) what the actual values of δ_{ij} for the displayed products are. These two types of information are revealed to the consumer after the discovery step when she sees the product list page. Nevertheless, prior to discovering a set of products, the consumer knows the distribution of δ_{ij} for each search route, $G_r(\delta)$, with $r = 1, ..., R$ denoting search routes.

After looking at a product list page, the consumer learns δ_{ij} for the displayed products, but still does not know her match value with the product, ϵ_{ij}. However, she knows the distribution of ϵ_{ij}, denoted by $F(\epsilon)$. The match value is revealed to the consumer via search, i.e., when she visits a product detail page and looks at photos, reads product reviews or product descriptions.

This model differs from most frameworks in the consumer search literature that assume that consumers know δ_{ij} and do not need to discover it, i.e., that consumers are aware of all products and their characteristics shown on the list page and only need to decide what and whether to search to learn the value of ϵ_{ij} (e.g., Kim, Albuquerque, and Bronnenberg 2010; Chen and Yao 2017; Ursu, Zhang, and Honka 2023).

Next, we describe how we parameterize δ_{ij}, the part of the utility coming from product characteristics the consumer observes on the list page:

$$\delta_{ij} = \mathbf{x}_{ij}\beta_i + \eta_{ij}. \tag{2}$$

The value δ_{ij} consists of the following: (i) a vector of product characteristics $\mathbf{x}_{ij}$ such as price, a promotion dummy, an indicator for whether the consumer searched products in our focal category in the previous week (to proxy for her interest in the category), and a variable indicating whether the consumer visited the sales page (for any product) in the previous week (as a measure of the consumer's price sensitivity), (ii) a vector of consumer preferences for these characteristics β_i, and (iii) consumer i's product-specific idiosyncratic preference η_{ij}, e.g., how much she likes the sandal design. The idiosyncratic preference η_{ij} is known by the consumer upon discovery and before search but unobserved by the researcher. Before discovering a product, the consumer has rational expectations for the distribution of δ_{ij}. We assume that G_r follows a normal distribution $N\left(\mu_{\delta_r}, \sigma_{\delta_r}\right)$ and obtain its moments from data.[8] Also, we assume the distributions of both η_{ij} and ϵ_{ij} are standard normal.

Consumer i discovers and searches products sequentially with a discovery cost of c_i^d and a search cost of c_i^s. Both types of costs are parameterized as exponential functions, i.e., $c_i^d = \exp\left(\gamma_i^d\right)$ and $c_i^s = \exp\left(\gamma_i^s\right)$, to ensure that they are positive.[9] Paying a cost c_i^d allows the consumer to discover a set of n_d products. In our empirical application, we set $n_d = 4$ since consumers see four products on the app screen. Paying a cost of c_i^s allows the consumer

[8]We implicitly assume here that the route-specific distributions of δ are stable. In other words, we assume that the distribution of δ, from which the products the consumer sees on a screen are drawn, is constant. To put it differently, we do not allow for the case that "better" (worse) products are shown to consumers earlier (later) in the discovery process. Recall that δ consists of observable product characteristics, consumer preferences, and η_{ij}. η_{ij} is an individual- and product-specific unobservable which follows a standard normal distribution. Consumer preferences are stable by assumption. Using our data, we tested whether products which appear on earlier screens have statistically significantly different observable characteristics than products that appear on later screens. We neither find significant differences for prices nor for price promotions. And lastly, this retailer is a relatively unsophisticated retailer. Thus, we conclude that the assumption of stable route-specific distributions of δ is appropriate for our data.

[9]Although search costs have been modeled as functions of product rankings in prior work (Ursu 2018), given that we reconstructed the product order in our data from click frequencies (see Section 3.1) and given that only four products are displayed on each screen, we do not take product ranking into account when parametrizing search costs and instead assume that search costs are consumer-specific.

to search one more product among those discovered, but not yet searched. Note that the consumer optimally chooses which product to search among those discovered (since products vary in their discovered attributes), as described in the next section.

Consumers are assumed to have perfect recall; they can search any discovered product and costlessly revisit any already-searched product. We also include an outside option (no purchase) in our model, i.e., consumers may not make a purchase at all (even after searching). For the outside option, we set $\delta_{i0} = 0$, i.e., $u_{i0} = \epsilon_{i0}$.

4.2 Optimal Consumer Behavior

Consumer i discovers and searches products sequentially. At every moment in time, she makes a choice between discovering more products, searching among the discovered products, or buying one of the searched products. The consumer can choose to discover or search in any order, except that she cannot search a product she has not discovered previously. Also, she cannot buy a product she has not searched. Finally, choosing to buy a searched product (including the outside option) terminates the process.

Greminger (2021) derived the rules governing optimal consumer behavior for the above problem. The rules involve three reservation utilities: the discovery reservation utility, z_{ir}^d, the search reservation utility, z_{ij}^s, and the purchase reservation utility, z_{ij}^b. The purchase reservation utility coincides with the utility of buying product j, so we can replace it with u_{ij}. The other two reservation utilities equate the marginal cost and the expected marginal benefit from each action, i.e., from discovering and searching.

The search reservation utility is defined as in Weitzman (1979):

$$c_i^s = \int_{z_{ij}^s}^{\infty} \left(u_{ij} - z_{ij}^s\right) f\left(u\right) du \ , \tag{3}$$

where c_i^s is the marginal cost of searching. The integral on the right-hand side gives the marginal benefit of one more search given that the best option revealed so far equals z_{ij}^s

and utility draws are distributed according to $F(u)$. In other words, the search reservation utility, z^s_{ij}, is defined as the utility realization of a product a consumer would have to have in hand in order to be indifferent between searching once more and stopping. Note that, although search costs are not product-specific, the observed utilities of discovered products vary across products affecting z^s_{ij}, which then affect whether and in what order a product will be searched.

Following Greminger (2021), the discovery reservation utility is given by

$$c_i^d = \int_{z_{ir}^d}^{\infty} \left[1 - H(w)\right] dw \ , \tag{4}$$

where c_i^d is the marginal cost of discovery. The integral on the right-hand side gives the marginal benefit of discovering more products when $H(\cdot)$ is the cumulative density of the maximum value of the δ values revealed in one discovery step. Note that z^d_{ir} is route- but not product-specific: choosing to discover products via route r reveals n_d products with values according to the distribution of route r.

Intuitively, optimal behavior is determined by the relative values of the three reservation utilities. That is, at a moment in time, if among the three reservation utilities, the discovery reservation utility is the largest, then the consumer will discover another set of products. If instead, the search reservation utility is the largest, then the consumer will search. Alternatively, the consumer will stop and make a purchase decision if the purchase reservation utility is the largest (see Greminger (2021) for a proof of this argument).

More formally, at every step $t \in \{1, 2, \ldots, T\}$, the consumer makes a choice between discovering more products, searching among the discovered products, or buying one of the searched products. Let S_{it} denote the set of products searched before step t and D_{it} denote the set of products discovered before step t. For example, S_{i1} consists of the outside option and D_{i1} consists of the products consumer i is aware of before starting the process, i.e., the first four products on the first route's list page. Paying a cost c_i^d allows the consumer to discover a new set of n_d products, while paying a cost c_i^s allows the consumer to search one more product among those discovered, but not yet searched. At t, the set of products available for discovery is given by $J \setminus D_{it}$, while the set of products available for search is given by $D_{it} \setminus S_{it}$. Finally, the set of products the consumer can purchase from is given by S_{it}.

For notational simplicity, we drop the consumer-specific subscript i in what follows. Further, we define the maximum search reservation value at t as $\tilde{z}^s(t) = \max_{j \in D_t \setminus S_t} z_j^s$ and the maximum purchase utility at t as $\tilde{u}(t) = \max_{j \in S_t} u_j$. Then, following Theorem 1 in Greminger (2021), the optimal search rules are given by[10]

1. *Stopping Rule:*

   Purchase product $j \in S_t$ and end search whenever $\tilde{u}(t) \geq \max\{z_r^d, \tilde{z}^s(t)\}, \forall r \in R$.

2. *Search Rule:*

   Search $j \in D_t \setminus S_t$ whenever $\tilde{z}^s(t) \geq \max\{z_r^d, \tilde{u}(t)\}, \forall r \in R$.

3. *Discovery Rule:*

   Discover more products whenever $\max\{z_r^d\} \geq \max\{\tilde{z}^s(t), \tilde{u}(t)\}, \forall r \in R$.

Once the search and discovery process ceases, the consumer chooses to buy the product with the largest realized utility among those searched (*Choice Rule*). These optimal search rules allow us to estimate consumer preferences, search costs, and discovery costs using our data. We provide details of our estimation approach next.

5 Estimation

5.1 Estimating Reservation Utilities

As shown in Section 4.2, optimal consumer behavior is dictated by a set of three reservation utilities. Before discussing how we estimate the model primitives, we need to describe how we

[10] To be consistent with prior empirical work on consumer search, we renamed the "inspection rule" in Greminger (2021) to be the "search rule."

estimate these reservation utilities. First, recall that the purchase reservation utility coincides with the purchase utility and is therefore estimated using the expression given in equation (1).

Second, following Kim, Albuquerque, and Bronnenberg (2010), we use the lookup method to back out search reservation utilities from equation (3). Given the standard normal assumption for ϵ_{ij}, it follows that the search reservation utilities have the form

$$z_{ij}^s = \delta_{ij} + g(c_i^s), \tag{5}$$

where $g(\cdot)$ is a known function that monotonically decreases in search costs (Kim, Albuquerque, and Bronnenberg 2010). We refer the reader to Ursu, Seiler, and Honka (2023) for details on the implementation of the lookup method.

Following Greminger (2021) and assuming that δ and ϵ are independent, the discovery reservation utility can be written as

$$z_{ir}^d = \mu_{\delta_r} + M_r(c_i^s, c_i^d, n_d, \sigma_{\delta_r}), \tag{6}$$

where $M(\cdot)$ solves equation (4) for the random variable $\tilde{\delta} = \delta - \mu_\delta$. Recall that μ_{δ_r} and σ_{δ_r} are the empirical mean and standard deviation of δ for route r.

Greminger (2021) provides expressions for the integral derived in equation (4) that allow a researcher to compute $M(\cdot)$ in equation(6). To develop intuition for these expressions, we use simulations to illustrate how $M(\cdot)$ varies with its arguments. For this illustration, we focus on one consumer (suppressing the subscript i in what follows) and compute the value of $M(\cdot)$ at fixed values of all its arguments, except one that we vary to highlight its effect on $M(\cdot)$. The values of $M(\cdot)$'s arguments when they are fixed are ($c_s = \exp(-2)$, $c_d = \exp(-1)$, $n_d = 4$, $\sigma_{\delta_r} = 1$).

The results are displayed in Figure 3. Both larger values of the search cost and of the discovery cost are associated with lower values of $M(\cdot)$. This relation is intuitive: if search or discovery costs are larger, the value of continued discovery is lower, as reflected in a lower $M(\cdot)$ and therefore in a lower discovery reservation value, z^d_{ir}. We also show that a larger n_d, i.e., more options discovered per cost paid, increases $M(\cdot)$ and z^d_{ir} since the benefit from a discovery decision increases. Finally, the larger the variance of options σ_{δ_r} revealed through product discovery, the larger $M(\cdot)$ and z^d_{ir}, reflecting the option value of additional discovery (with similar results shown in prior work for the relation between variance and search reservation utilities; see Weitzman 1979).

==========================

Insert Figure 3 about here

==========================

We contribute to the Greminger framework by using equation (4) and the lookup table method to back out discovery and search reservation utilities from equation (6) directly. In contrast, Greminger (2022) specifies an expression for both the discovery and search reservation utilities and then proceeds to estimate parameters of this expression. This method is fast, but does not jointly and directly estimate search and discovery costs (which are not needed in Greminger 2022). Also, in our empirical application, we allow consumers to discover products via multiple routes, rather than using only a single route.

5.2 Likelihood Function

We use the search rules described in Section 4.2 to construct the likelihood of observing a certain sequence of discovery, search, and purchase decisions in our data. These rules translate into the following restrictions on preferences, search costs, and discovery costs. At every step t, the consumer decides whether to discover, search, or terminate the process by making a purchase decision in some final period T. Let once again S_{it} denote the set of products searched before step t and D_{it} denote the set of products discovered before step t.

A consumer discovers more products at step $t < T$ if the discovery reservation utility of at least one route r exceeds all search reservation utilities (among the already discovered but not yet searched products) and all purchase utilities (among the searched products). Thus, according to the discovery rule, if we observe a discovery decision at t, then it must be that

$$\max\{z_r^d\} \geq \max\{z_j^s, u_k\} \qquad \forall r \in R, \forall j \in \{D_t \setminus S_t\}, \forall k \in S_t, \forall t < T. \tag{7}$$

A consumer searches product j at step $t < T$, if she has not searched it yet, if it has the largest search reservation utility among discovered but not yet searched products, and if its search reservation utility is larger than the maximum realized utility so far and the maximum discovery reservation utility. Formally, according to the search rule, the consumer searches j if

$$\max\{z_j^s\} \geq \max\{z_r^d, u_k\} \qquad \forall r \in R, \forall j \in \{D_t \setminus S_t\}, \forall k \in S_t, \forall t < T. \tag{8}$$

According to the stopping rule, consumer stops searching at step $t = T$ when the maximum realized utility among the searched products is larger than the maximum of the discovery reservation and search reservation utilities, i.e.,

$$\max\{u_k\} \geq \max\{z_r^d, z_j^s\} \qquad \forall r \in R, \forall j \in \{D_T \setminus S_T\}, \forall k \in S_T. \tag{9}$$

Finally, once the consumer stops searching, she purchases the alternative with the highest realized utility among the searched ones and the outside option, i.e.,

$$u_j^* \geq \max\{u_k\} \qquad \forall k \in S_T. \tag{10}$$

Although Greminger (2021) does not call it that, consistent with prior work on search, we refer to the last rule as the "choice rule" (e.g., Kim, Albuquerque, and Bronnenberg 2010; Honka 2014; Ursu 2018; Ursu, Seiler, and Honka 2023).

If the consumer behaves using the rules described above, then she makes discovery, search, and purchase decisions jointly. Therefore, the probability of observing a certain outcome in the data for a given consumer is characterized by the joint probability of the choice, stopping, search, and discovery rules holding at every step, i.e., by

$$L_i = Pr\,(\text{ Discovery rule, Search rule, Stopping rule, Choice rule })\,. \tag{11}$$

The individual likelihood function cannot be expressed in closed form. We approximate the integrals in the likelihood function with averages using logit-smoothed accept-reject simulation. This simulated maximum likelihood estimation (SMLE) algorithm follows Train (2009) and is widely used in the search literature (e.g., Honka 2014; Honka and Chintagunta 2017; Ursu 2018; Ursu, Zhang, and Honka 2023). Implementation details are discussed in Web Appendix B.

We first estimate our model at the session level, capturing all discovery, search, and purchase decisions consumers make before deciding to end the session. Recall that our data has a panel component, i.e., we observe consumers searching across multiple sessions. Thus, we also estimate our model allowing for unobserved heterogeneity in preferences and costs by estimating a two latent segments model (Dayton and Macready 1988; Kamakura and Russell 1989). In the two latent segments model, we estimate the probability of a consumer belonging to segment 1 as $\pi_1 = \exp(\rho)/(1+\exp(\rho))$ and the probability of a consumer belonging to the segment 2 as $\pi_2 = 1 - \pi_1$. In all estimations, we normalize the price variable by dividing it by its maximum value to make its scale comparable to that of the other variables. Furthermore, our estimation sample focuses on consumers who search via category page and/or sales pages. The estimation sample contains 1,342 consumers searching across 2,109 sessions. More details on the considerations that led to this estimation sample can be found in Web Appendix B.

5.3 Identification

The parameters to be estimated include the preference parameters β_i, the discovery cost c_i^d (parameterized by γ_i^d), and the search cost c_i^s (parameterized by γ_i^s). Here, we present an informal discussion of identification. For a formal discussion, we refer the reader to Ursu, Seiler, and Honka (2023). The preference parameters β are identified by purchase frequency, search order, and search frequency. In much the same way that the purchase decision among a set of products identifies preference parameters in a traditional discrete choice model, the purchase decision among searched products identifies preference parameters in our model. Also, products that are searched more frequently and those that are searched more often first will have a larger estimated β value. Finally, choosing to discover products from one route more frequently will indicate higher preference for certain product attributes that distinguish that route from others (e.g., higher price sensitivity when choosing to search through sales pages).

Both search and discovery costs do not affect purchase decisions. They are identified from the number of products consumers discover and search, respectively. Holding everything else constant, consumers search little when c_i^s is large and a lot when c_i^s is small. A similar pattern holds for c_i^d. In addition, search costs affect the value of the discovery reservation utilities, and therefore affect discovery decisions. Higher search costs imply higher discovery costs, discouraging further discovery. Differences in functional form between the direct effect of search costs on search reservation utilities versus through discovery reservation utilities allow us to separately pin down these values.

5.4 Monte Carlo Simulation Study

We perform the following Monte Carlo simulation exercise to show that our estimation procedure recovers the true parameters. We generate a data set of 2,000 consumers making discovery, search, and purchase decisions using one search route. There are 1,000 products available, but only a set of 10 (random) products are available to each consumer (to keep the simulated data from increasing too fast in size). This specification allows us to keep the data manageable in size, as well as mirrors the method used by Greminger (2021). Consumer utility has three components corresponding to the three possible observables in our data set: price, a promotion dummy, and an indicator for previous searches. These observables are constructed with mean and standard deviation matching our data in Table 1 (except that we use the log of the price and normalize it by dividing by its maximum to ensure that all variables are on the same scale). Consumers incur a cost to discover two products at a time (we let $n_d = 2$ in the simulation given that there are 10 products available to each consumer) and a cost to search a product they have already discovered. Consumers begin the process with one (random) product they are aware of (one product in the discovery set) and an outside option they can choose if they decide not to buy anything (which enters the initial searched set). The true values of the three utility and two cost parameters are relatively similar to those from a preliminary estimation of our model.

For estimation, we follow the steps described in the previous section and use 200 draws from the distributions of the utility error terms (both η_{ij} and ϵ_{ij}) for each consumer-product combination. We simulate 100 different data sets using the same true parameters but different seeds for the utility error terms and repeat the estimation for each data set.

Our Monte Carlo simulation results are displayed in Table 3. In column (i), we present the true parameters; in column (ii), we show the mean of the estimated parameters across the 100 simulations and the standard deviation of the mean across these simulations. As can be seen, our proposed estimation procedure recovers the model parameters well.

==========================

Insert Table 3 about here

==========================

6 Results

6.1 Model Estimates

In Table 4, we report the estimation results for our model both when assuming one latent consumer segment (column (i)), i.e., homogenous consumer preferences as well as homogenous search and discovery costs, and when considering the case of two latent segments (column (iii)). Finally, we also contrast our results with those obtained if we ignored consumers' discovery decisions and instead estimated the Weitzman (1979) model on the same data (column (ii)).

==========================

Insert Table 4 about here

==========================

We start by discussing the estimation results for our model with homogenous preferences, search and discovery costs (column (i)). As expected, the price coefficient is negative, whereas the promotion dummy is positive, i.e., consumers are more likely to search and purchase a product that is inexpensive and/or that is being promoted. Visiting the sales page in the week prior to the current search also shows a significant positive effect, suggesting that such consumers are more likely to search and buy products in the current session. Both search costs and discovery costs are statistically significant. In terms of magnitude, search costs ($\exp(0.18) = 1.19$) are approximately four times larger than discovery costs ($\exp(-1.31) = 0.27$). This result explains why consumers frequently search only a subset of the discovered products.

In column (ii), we present the results from the Weitzman (1979) model which assumes that consumers are aware of all products the retailer sells, i.e., assumes there is no need to discover products. While the signs and significance levels of all but one estimated coefficient in the Weitzman (1979) model are similar to those in our model, the magnitudes of the coefficients are different. For example, the estimated price and promotion coefficients are much larger (in absolute terms). Also, visiting the sales page in the week prior to the current search has a negative effect on consumer utility (rather than a positive effect as in our model). Finally, we obtain a search cost that is three times smaller than in our model ($\exp(-0.94) = 0.39$).

The difference in estimated parameters comes from the different assumptions made on product discovery in the two models. In our model, prices and other attributes of products that are on product list pages the consumer never sees/discovers do not affect choices. For example, if higher priced options are available but never discovered, then the Weitzman (1979) model, which assumes that all products were discovered, will estimate a higher price sensitivity than our model, as we report in Table 4. In addition, by assuming that consumers are aware of all available products (essentially having zero discovery cost), the Weitzman (1979) model underestimates search costs, as they are confounded with low discovery costs. In contrast, our model is able to separate search and discovery decisions, revealing a higher search cost and low discovery costs. Given these differences, we find that the Weitzman (1979) model fits the data worse than our model.

Finally, in column (iii) we present results from our model with two latent consumer segments. We find that segment 1 is relatively larger than segment 2 (segment 1 accounts for 63% of the sample)[11] and that the two segments vary in their parameter estimates. More precisely, consumers in segment 1 are directionally less sensitive to regular prices than consumers in segment 2 (though the difference is not statistically significant), but more sensitive to a pair of sandals being promoted. While having searched for sandals in the previous week has no significant effect on either segment of consumers, having visited the app's sale page (for products of any category) has significant effects on both segments of consumers with the coefficient being larger for consumers belonging to segment 1. Further, consumers in segment 1 have lower discovery and search costs than consumers in segment 2. Taken together, we interpret these results as indicating that consumers in segment 1 are individuals who seek promotions in general as well as for sandals in particular and respond more strongly when they find one than consumers in segment 2.

[11]Computation obtained since the probability of segment 1 is given by $\pi_1 = \exp(\rho)/(1+\exp(\rho)) = \exp(0.55)/(1+\exp(0.55)) = 0.63$.

6.2 Price Elasticity

To calculate the price elasticity, we first simulate discovery, search, and purchase behavior from the fitted 1-segment model. We then decrease the prices for all products by 10% and re-simulate consumer behavior. Note that we decrease the prices observed by the consumer after discovery, but keep the price distribution unchanged. We repeat this exercise 30 times for each session in order to integrated out over the distribution of unobserved utility error terms. The price elasticity is computed as the average difference in percent between the simulated outcomes with and without a 10% price decrease. The price elasticity is estimated to be -0.114 for discovery, 0.017 for search, and 0.041 for purchase. In other words, a price reduction increases search and purchase probabilities. In addition, lowering the price also affects product discovery because consumers are more likely to find a product they want to purchase earlier and therefore terminate the process earlier than without the 10% price reduction.

The price elasticity discussed in the previous paragraph is the elasticity for the regular (unpromoted) price. The actual price consumers pay is a combination of the regular price and potentially a price promotion. Recall that we only observe whether a product is being promoted but not the promotional depth. To give a more complete picture of consumers' response to price changes, we also predicted consumer behavior when the promotional dummy switches from 0 to 1 for all available products. In such a case, discovery decreases by 25.67%, but searches and purchases increase by 5.93% and 16.51%, respectively. The negative effect on product discovery is, again, driven by consumers finding a product they want to purchase earlier in the presence of promoted products.

6.3 Discovery and Search Cost Elasticities

To calculate the discovery cost and search cost elasticities, we follow the same procedure as outlined in the previous subsection for the price elasticity. We predict how consumer discovery, search, and purchase behavior changes due to 20% and 50% changes in each cost type.

Recall that search costs do not only affect search reservation utilities, but also discovery reservation utilities (see equation (6) and Figure 3). In other words, if search costs are large, then the benefit from searching is small and, at the same time, the benefit from discovery is low since consumers will not be able to search many of the discovered products. Our results in Table 4 reveal that consumers have search costs that are four times larger than discovery costs. Therefore, decreasing discovery costs or increasing search costs even further can lead to corner effects on consumer decisions. For these reasons, we report discovery and search cost elasticities as a result of an increase in discovery costs and a decrease in search costs.

We show our results in Table 6. In discussing the elasticities, we focus on the results from a 50% change in each type of cost, given the similarity in the direction of the results. Not surprisingly, consumers discover 10.17% fewer products and search 1.53% fewer products when discovery costs are increased. This lower discovery and search activity results in 3.49% fewer product purchases. In contrast, lowering search costs leads to more searches and more purchases: 13.68% more products are searched and 19.63% more products are purchased. Interestingly, a lower search cost decreases the number of products discovered. Therefore, even though lower search costs encourage more discovery, they also encourage more search, leading to an overall negative effect on discovery for our model estimates.

==========================

Insert Table 6 about here

==========================

7 Counterfactuals

In this section, we conduct two counterfactuals: we measure the (marginal) value of a search route to the platform and we investigate the demand-side consequences of redesigning the product list page to make fewer or more products visible on a screen without needing to scroll down. We note that neither counterfactual could be conducted within the Weitzman (1979) framework since neither product discovery nor search routes are part of the model.

The counterfactuals are implemented as follows: we first simulate discovery, search, and purchase behavior using our preference, search and discover cost estimates. Then, we impose the counterfactual changes and again simulate discovery, search, and purchase behavior for these alternative scenarios. And lastly, we compare simulated behavior under these changes with behavior without such changes. We repeat this exercise 30 times for each session to integrate over the utility error draws and report the average difference in percent between the simulated outcomes.

7.1 The Value of a Search Route

In the first counterfactual, we measure the (marginal) value of a search route to the platform by investigating how consumer shopping behavior changes when a search route is removed. Specifically, we focus on the effect of removing the "Sales page" route. Given the prominence and importance of the main "Category page" on the website, we do not view its removal as a managerially relevant scenario. Further, since our counterfactuals are only able to capture behavior changes that are not too far from the original environment, they may not be able to fully capture the effects of such a disruptive change in the app design.

To perform this counterfactual, we assume that consumers are only allowed to discover, search, and purchase products via the "Category page" route, i.e., they cannot choose between two routes. We find that removing the option of using the "Sales page" route decreases product discovery by 11.51%, product search by 1.72%, and product purchases by 3.98%, i.e., consumers discover, search, and purchase fewer sandals. In other words, consumers do not "simply" switch to a different search route and purchase as before; removing a search route has a negative effect on purchase behavior. Finally, the decline in searches and purchases is much smaller than the decline in discovery.

As mentioned, purchases decline by about 4% when the sales route is removed. However, that does not necessarily mean that the online retailer is worse off. For example, if consumers were to purchase more expensive sandals using the category page, that could potentially offset the decline in purchase probability. However, we find that the online retailer's revenue decreases by 4.2%, i.e., the decline in revenue in percent is even somewhat larger than the decline in purchase probability. The 4.2% of the revenue also represents the marginal value of the search route to the online retailer.

7.2 Screen Design

In our second counterfactual, we examine the effects of changing the number of products visible on a screen without needing to scroll down. Recall that the mobile app currently shows four products to the consumer on each product list page. Here, we quantify the consequences for product discovery, search, and purchase of showing more or fewer products on each page. By changing the number of visible products, the app can affect the number of products consumers can discover with every such decision. We note that we only perform this counterfactual around the current number of products shown on a screen, since increasing the number of products to a much larger number could introduce considerations (e.g., information overload) that our model does not account currently for.

Our results are displayed in Table 7. When the mobile app switches from four to two products, i.e., shows fewer products on each screen, consumers discover 28.02% fewer products, search 3.43% fewer products, and buy 8.42% fewer products. In contrast, when the mobile app increases the number of products shown on a screen from four to six, consumers discover 7.62% more sandals, search about the same number of sandals, and purchase 0.15% more sandals. Therefore, by changing the number of discovered and then searched products, this new design can affect consumer final purchases. Since increasing the number of products visible on a screen to six likely reduces the readability (which our model does not account for), while only marginally increasing sales, we interpret these results as evidence that displaying four products on a screen may be optimal for the retailer.

==========================

Insert Table 7 about here

==========================

8 Conclusion

In this paper, we use data from the mobile app of an apparel retailer and document that consumers frequently use different routes to reach the same products. To model consumers' decisions of not only what to search and buy, but also of the route through which to discover products, we follow Greminger (2021, 2022) and estimate a model of joint discovery, search, and purchase decisions. We use our model to quantify preferences, discovery costs, and search costs, and show that product search costs are four times larger than product discovery costs. Finally, we quantify the (marginal) value of a search route to the online retailer and illustrate the impact of an app design change on product discovery and sales via counterfactuals.

Our research has several limitations that may lead to useful extensions. First, in our model, as well as in Greminger (2021, 2022), consumers' beliefs about the value of using a given search route are independent of other routes used. Also, such beliefs are constant throughout the search process, i.e. are not affected by the utilities of the products searched and discovered so far. However, it is possible that such expectations may not be independent or may not remain constant while searching. For example, if the app ranks products in decreasing order of attractiveness and consumers know this, then they may become increasingly more pessimistic as they discover and search further. This would lead them to terminate the process earlier than in the current model. We leave these and other related topics to future research with data on the specific algorithm the platform used or on consumers' changing beliefs. Second, we only have data from a mobile commerce retailer. We are thus not able to compare the search cost between traditional e-commerce and mobile commerce. Future research can obtain data from retailers who operate in both channels and compare consumer discovery, search, and purchase behavior. And finally, with data on products displayed from more search routes, future work can provide a more detailed picture of the entire consumer shopping process.

References

Bairathi, Mimansa, Xu Zhang, and Anja Lambrecht (2022), "The Value of Platform Endorsement," *Working Paper*.

Bang, Youngsok, Dong-Joo Lee, Kunsoo Han, Minha Hwang, and Jae-Hyeon Ahn (2013), "Channel capabilities, product characteristics, and the impacts of mobile channel introduction," *Journal of Management Information Systems*, 30 (2), 101–126.

Bronnenberg, Bart J, Jun B Kim, and Carl F Mela (2016), "Zooming in on Choice: How do Consumers Search for Cameras Online?" *Marketing Science*, 35 (5), 693–712.

Chen, Pei-Yu and Lorin M Hitt (2002), "Measuring Switching Costs and the Determinants of Customer Retention in Internet-enabled Businesses: A Study of the Online Brokerage Industry," *Information systems research*, 13 (3), 255–274.

Chen, Yuxin and Song Yao (2017), "Sequential Search with Refinement: Model and Application with Click-Stream Data," *Management Science*, 63 (12), 4345–4365.

Dayton, C. and G. Macready (1988), "Concomitant-variable Latent-class Models," *Journal of the American Statistical Association*, 83 (401), 173–178.

De los Santos, Babur, Ali Hortaçsu, and Matthijs R Wildenbeest (2012), "Testing Models of Consumer Search Using Data on Web Browsing and Purchasing Behavior," *American Economic Review*, 102 (6), 2955–80.

Fang, Zheng, Bin Gu, Xueming Luo, and Yunjie Xu (2015), "Contemporaneous and Delayed Sales Impact of Location-based Mobile Promotions," *Information Systems Research*, 26 (3), 552–564.

Gardete, Pedro and Megan Anthill (2020), "Guiding Consumers through Lemons and Peaches: An Analysis of the Effects of Search Design Activities," *Working Paper*.

Ghose, Anindya, Avi Goldfarb, and Sang Pil Han (2013), "How is the Mobile Internet Different? Search Costs and Local Activities," *Information Systems Research*, 24 (3), 613–631.

Greminger, Rafael (2021), "Optimal Search and Discovery," *Management Science*, 68 (5), 3904–3924.

— (2022), "Heterogeneous Position Effects and the Power of Rankings," *Working Paper*.

Hodgson, Charles and Gregory Lewis (2021), "You Can Lead a Horse to Water: Spatial Learning and Path Dependence in Search Data," *Working Paper*.

Honka, Elisabeth (2014), "Quantifying Search and Switching Costs in the U.S. Auto Insurance Industry," *The RAND Journal of Economics*, 45 (4), 847–884.

Honka, Elisabeth and Pradeep Chintagunta (2017), "Simultaneous or Sequential? Search Strategies in the U.S. Auto Insurance Industry," *Marketing Science*, 36 (1), 21–42.

Honka, Elisabeth, Ali Hortaçsu, and Maria Ana Vitorino (2017), "Advertising, Consumer Awareness, and Choice: Evidence from the U.S. Banking Industry," *The RAND Journal of Economics*, 48 (3), 611–646.

Kamakura, Wagner and Gary Russell (1989), "A Probabilistic Choice Model for Market Segmentation and Elasticity Structure," *Journal of Marketing Research*, 26 (4), 379–390.

Kim, Jun B, Paulo Albuquerque, and Bart J Bronnenberg (2010), "Online Demand Under Limited Consumer Search," *Marketing Science*, 29 (6), 1001–1023.

— (2017), "The Probit Choice Model under Sequential Search with an Application to Online Retailing," *Management Science*, 63 (11), 3911–3929.

Koulayev, Sergei (2014), "Search for Differentiated Products: Identification and Estimation," *The RAND Journal of Economics*, 45 (3), 553–575.

Morozov, Ilya (2023), "Measuring Benefits from New Products in Markets with Information Frictions," *Management Science*, forthcoming.

Morozov, Ilya, Stephan Seiler, Xiaojing Dong, and Liwen Hou (2021), "Estimation of Preference Heterogeneity in Markets with Costly Search," *Marketing Science*, 40 (5), 871–899.

Ngwe, Donald, Kris Johnson Ferreira, and Thales Teixeira (2019), "The Impact of Increasing Search Frictions on Online Shopping Behavior: Evidence from a Field Experiment," *Journal of Marketing Research*, 56 (6), 944–959.

Train, Kenneth E (2009), *Discrete Choice Methods with Simulation*, Cambridge University Press.

Ursu, Raluca (2018), "The Power of Rankings: Quantifying the Effect of Rankings on Online Consumer Search and Purchase Decisions," *Marketing Science*, 37 (4), 530–552.

Ursu, Raluca, Stephan Seiler, and Elisabeth Honka (2023), "The Sequential Search Model: A Framework for Empirical Research," *Working Paper*.

Ursu, Raluca, Qingliang Wang, and Pradeep Chintagunta (2020), "Search Duration," *Marketing Science*, 39 (5), 849–871.

Ursu, Raluca, Qianyun Zhang, and Elisabeth Honka (2023), "Search Gaps and Consumer Fatigue," *Marketing Science*, 42 (1), 110–136.

Venkatesh, Viswanath and Ritu Agarwal (2006), "Turning Visitors into Customers: A Usability-centric Perspective on Purchase Behavior in Electronic Channels," *Management Science*, 52 (3), 367–382.

Wang, Rebecca Jen-Hui, Edward C Malthouse, and Lakshman Krishnamurthi (2015), "On the Go: How Mobile Shopping Affects Customer Purchase Behavior," *Journal of Retailing*, 91 (2), 217–234.

Weitzman, Martin (1979), "Optimal Search for the Best Alternative," *Econometrica*, 641–654.

Wells, John D, Joseph S Valacich, and Traci J Hess (2011), "What Signal are you Sending? How Website Quality Influences Perceptions of Product Quality and Purchase Intentions," *MIS Quarterly*, 35 (2), 373–396.

Xu, Kaiquan, Jason Chan, Anindya Ghose, and Sang Pil Han (2017), "Battle of the Channels: The Impact of Tablets on Digital Commerce," *Management Science*, 63 (5), 1469–1492.

Zhang, Shuo, Zhenling Jiang, and Hai Che (2019), "Online Shopping with Endogenous Channel Choice: PC vs. Mobile," *Working Paper*.

Zhang, Xingyue, Ruomeng Cui, and Oliver Yao (2023), "The Version Effect of Apps and Operating Systems in Mobile Commerce," *Production and Operations Management*, 32 (2), 637–654.

Figures and Tables

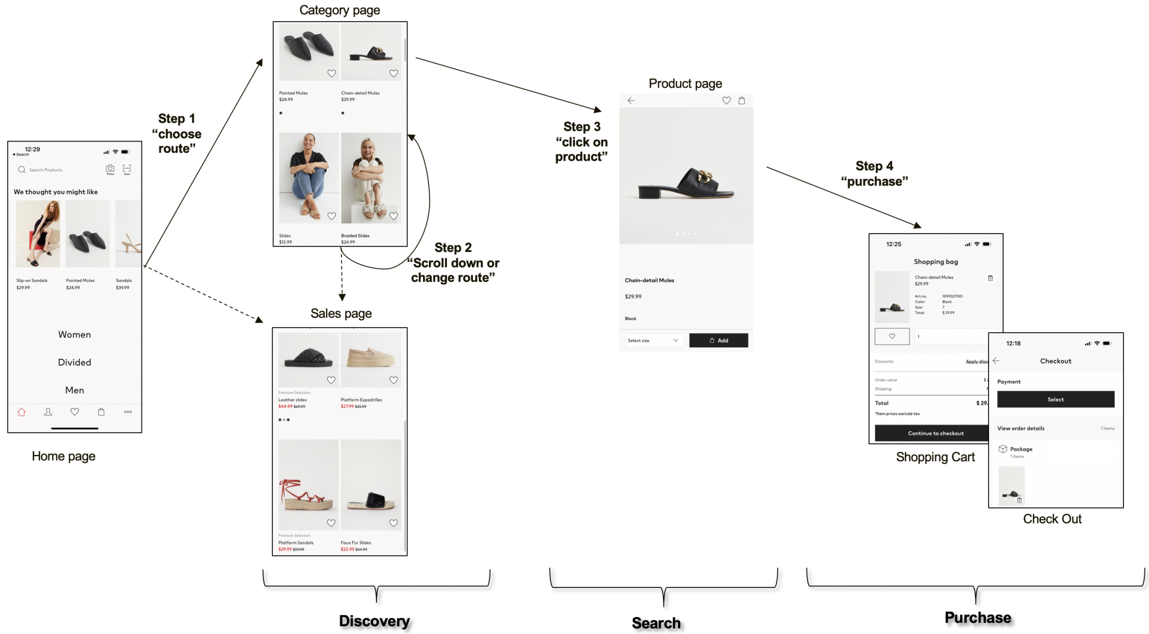
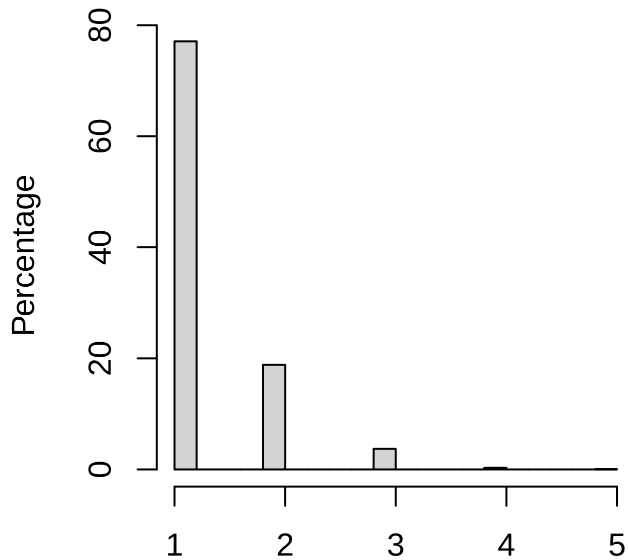
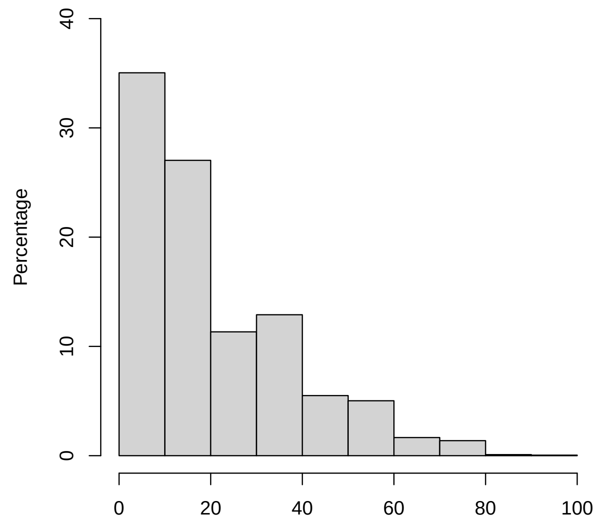


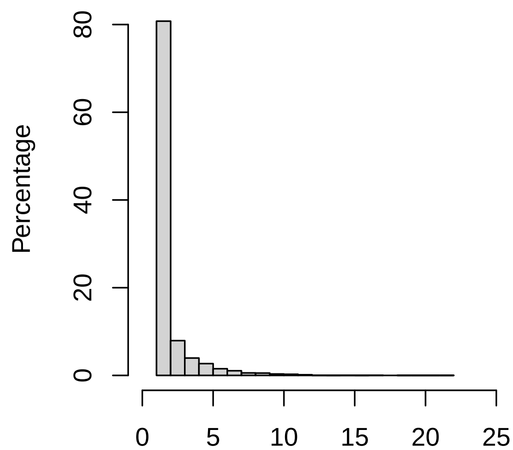
Figure 1: Example of Consumer Decision Sequence



(a) Number of Routes



(b) Number of Discovered Products



(c) Number of Searched Products

Figure 2: Histograms of the Number of Routes, Discovered Products, and Searched Products per Session

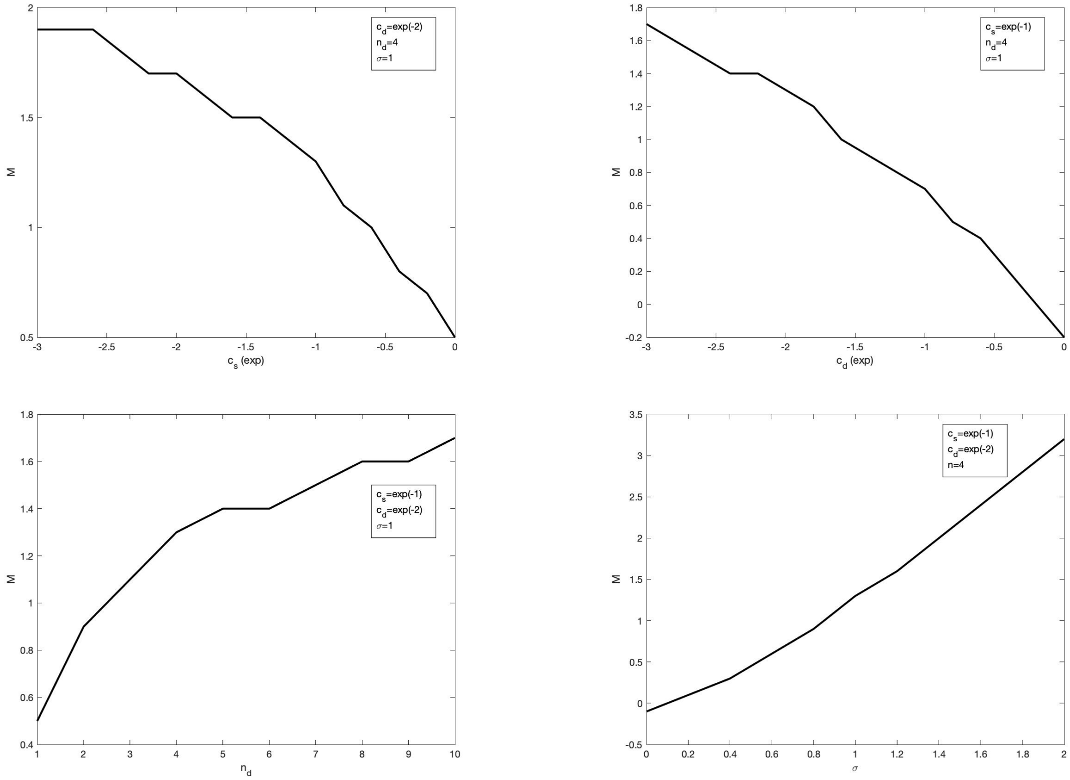


Figure 3: Value of $M(\cdot)$ as a Function of its Components

Variables	Mean	Std. Dev.	Min	Median	Max
Number of Routes per Session	1.27	0.54	1	1	5
Number of Searches per Session	1.88	1.76	1	1	22
Number of Discovered Products per Session	21.22	5.31	4	16	96
Available Products					
Price	37.28	11.94	7.99	37.99	75.00
Promotion Indicator	0.02				
Clicked Products					
Price	29.10	11.34	7.99	27.83	75.00
Promotion Indicator	0.14				
Consumer-Session Characteristics					
Searched Sandals Past Week	0.36				
Searched on Sales Page Past Week	0.36				

Table 1: Descriptive Statistics

	Av. Number of Searched Products	Av. Proportion of Promoted Product	Av. Price of Searched Products	Purchase Probability	Av. Price of Purchased Product	N. of Obs (Number of Sessions)
Homepage	1.11	26%	26.17	1%	26.99	799
Search Routes						
Category	2.15	1%	33.83	4%	25.95	799
Sales	1.18	100%	26.41	1%	25.10	2,192
Search Function	1.62	2%	32.20	8%	29.82	367
Recommendations	2.15	1%	31.24	2%	25.62	2,394
Product Page	1.32	1%	28.84	3%	24.53	1,608
Other	1.20	4%	28.95	3%	27.24	1,156

Notes: "Homepage" is the home screen that consumers see when they open the mobile app. "Category" is the category page. "Sales" is the sales page. "Search Function" describes the route used to look for a specific term by typing it into the search box and subsequently seeing a list page matching this search query. "Recommendations" relates to discovering products via recommendations available on the home page, on product pages, etc. "Product Page" describes the situation in which a consumer directly moves from a product page to another product page (not based on a recommendation). "Other" contains all other routes, such as discovering products through the consumer's account page, on the shopping cart page, etc.

Table 2: Descriptive Statistics for Search Routes

	(i) True values	(ii) Estimates	Std. Dev.
Utility			
X_1	-0.55	-0.48	(0.15)
X_2	2.00	2.17	(0.20)
X_3	1.00	1.36	(0.25)
Cost (exp)			
Discovery	-2.0	-1.72	(0.33)
Search	-2.0	-2.20	(0.75)
Log-likelihood		-4,495	
Number of Consumers		2,000	

Notes: Data are simulated for 2,000 consumers and the reported results are obtained after averaging across 100 estimations with different seeds and with 200 error draws each. The standard deviation of the mean estimate across these simulations is reported in parentheses.

Table 3: Monte Carlo Simulation Results

	(i) 1-Segment Model		(ii) Weitzman (1979)		(iii) 2-Segment Model			
	Coefficients	Std. Errors	Coefficients	Std. Errors	Coefficients	Std. Errors	Coefficients	Std. Errors
Utility								
Price	-0.48**	(0.16)	-1.63***	(0.11)	-0.55***	(0.14)	-0.70**	(0.23)
Promotion	0.53***	(0.06)	1.31***	(0.04)	0.55***	(0.07)	0.43***	(0.11)
Searched Sandals Past Week	0.07	(0.13)	-0.07	(0.09)	0.11	(0.20)	0.09	(0.22)
Searched on Sales Page Past Week	0.41***	(0.08)	-1.95***	(0.08)	0.53***	(0.09)	0.27**	(0.09)
Segment 1 Probability $\pi_1 = \exp(\rho)/(1+exp(\rho))$								
ρ					0.55*	(0.25)		
Cost (exp)								
Discovery	-1.31***	(0.00)			-1.31***	(0.00)	-1.00***	(0.15)
Search	0.18***	(0.04)	-0.94***	(0.07)	0.15**	(0.06)	0.26***	(0.05)
Number of Sessions	2,109		2,109		2,109			
Number of Consumers	1,342		1,342		1,342			
Log-Likelihood	-5,636		-8,144		-5,659			
BIC	11,310		16,327		11,396			

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

Notes: Reported results are obtained after averaging across 30 estimations with different seeds and with 50 error draws each.

Table 4: Estimation Results

	Price Reduction by 10%	Promotion Turned On/Off
Discovery	-1.14	-25.67
Search	0.17	5.93
Purchase	0.41	16.51

Notes: Percent change in the number of products searched, discovered and purchased in each session. To compute these changes, we simulate choices based on our estimates and compare them to the case where price is reduced by 10% or promotions are all turned on rather than off. We perform the simulation from 30 different seeds and present average results in the table.

Table 5: Price and Promotion Elasticities in %

	Discovery Cost Increase by		Search Cost Reduction by	
	20%	50%	20%	50%
Discovery	-4.61	-10.17	-27.02	-25.63
Search	-0.71	-1.53	1.55	13.68
Purchase	-1.53	-3.49	1.52	19.63

Notes: Percent change in the number of products searched, discovered and purchased in each session. To compute these changes, we simulate choices based on our estimates and compare them to the case where discovery or search costs are changed. We perform the simulation from 30 different seeds and present average results in the table.

Table 6: Discovery and Search Cost Elasticities in %

	Number of Products on a Screen	
	Fewer (2)	More (6)
Discovery	-28.02	7.62
Search	-3.43	-0.11
Purchase	-8.42	0.15

Notes: Percent change in the number of products searched, discovered and purchased in each session. To compute these changes, we simulate choices based on our estimates and compare them to the case where consumers discover fewer or more products compared to the current default number, 4. We perform the simulation from 30 different seeds and present average results in the table.

Table 7: Changing the Number of Products Visible on a Screen in %

Web Appendix A: Data Cleaning and Reconstruction

Our data contain all click-stream activities of 10,000 randomly selected consumers on the retailer's mobile app between September 2017 and February 2018 (six months). These consumers can be located in any of the 90 countries the retailer operates in. We take the following steps to arrive at the data sample we use for the empirical analysis.

First, consumers who used more than one language on the shopping app, used multiple devices or appeared in several countries were dropped. Second, we restrict our data to consumers who clicked on a women's sandal product page (the largest product category) at least once and live in the nine countries with the most observations: Saudi Arabia, United Arab Emirates, Indonesia, Kuwait, Jordan, Egypt, Oman, Qatar, and the Kingdom of Bahrain. These two steps leave us with 5,275 consumers.

Third, following industry practice, we define search sessions as follows: a search session ends after 30 minutes of inactivity.[12] We drop consumers with sessions longer than 240 minutes. This step leaves us with information on 5,214 consumers and 16,972 sessions.

Fourth, we drop sandals searched fewer than three times over the 6-month time period, sandals which cost more than 75 in the local currency (about US $20)[13], and sessions in which a consumer did not browse at least one pair of women's sandals (after the aforementioned changes). If a consumer browsed any of the sandals we dropped in this step, we drop this consumer from our data. This step leaves us with information on 4,296 consumers and 10,687 sessions.

Fifth, women's sandals come in different subcategories, e.g., flat sandals, wedge sandals, etc., and we remove sessions in which a consumer searched in more than four sandal subcategories.

And lastly, we exclude clicks on administrative pages such as account setting pages,

[12] https://support.google.com/analytics/answer/2731565?hl=en#zippy=%2Cin-this-article

[13] Such prices are more than two standard deviations away from the average price in our data (see Table 1 for more details).

aggregate multiple clicks on the same product into one, and disregard sessions with multi-product purchases.[14]

We are left with 3,621 consumers browsing in 9,128 sessions, among which 203 sessions result in a purchase. We will use this data sample for our descriptive analysis. For our estimation, we focus on consumers who used the category page and/or sales pages as their search route(s).[15] This smaller sample contains 1,342 consumers and 2,109 sessions.

Although we do not observe products the consumer saw but did not click on, we utilize the three features of the retailer's app described in Section 3.1 to reconstruct the available products for consumers. Using the larger data set of 3,621 consumers, we sort searched products in the same search route during the same week by their search frequencies in descending order. The searched products consist of available products in a given route and week. We treat the descending order of searched frequency as the product ranking. Note that the mobile app shows four products per page. It is possible that the number of constructed available products in a route is not a multiple of four. In this case, we append the most searched products in routes other than the sales route or the category route until the number of available products is a multiple of four.

[14]Transaction data are not available. We impute purchase information from shopping cart additions and checkout information.

[15]There are two main data limitations: first, for some of the search routes, we do not observe and cannot reliably reconstruct the discovered but not searched products. This is the case for the search function and recommendations routes. And second, for some of the search routes, we do not have enough data to reliably estimate the empirical moments of the product distributions, as is the case for the goods detail page and the featured routes. These considerations lead us to focus on the category page and sales pages for our estimation.

Web Appendix B: Estimation Details

In this appendix, we provide more details on the simulated maximum likelihood estimation (SMLE) approach we use to infer the parameters of our model (following Train 2009; Honka 2014; Ursu 2018). This approach involves the following steps (we suppress the consumer i subscript in what follows):

1. Make $n = \{1, \ldots, N\}$ draws of η_j and ϵ_j for each consumer-product combination and compute $(u_j(n), z_j^s(n), z_r^d(n))$ as well as the discovery and search costs using model parameters.
2. Calculate the following expressions for each draw n:
   (a) $\nu_1(n) = \max\{z_r^d(n)\} - \max\{z_j^s(n), u_k(n)\} \quad \forall r \in R, \forall j \in \{D_t \backslash S_t\}, \forall k \in S_t, \forall t < T$
   (b) $\nu_2(n) = \max\{z_j^s(n)\} - \max\{z_r^d(n), u_k(n)\} \quad \forall r \in R, \forall j \in \{D_t \backslash S_t\}, \forall k \in S_t, \forall t < T$
   (c) $\nu_3(n) = \max\{u_k(n)\} - \max\{z_r^d(n), z_j^s(n)\} \quad \forall r \in R, \forall j \in \{D_T \setminus S_T\}, \forall k \in S_T$
   (d) $\nu_4(n) = u_j^*(n) \geq \max\{u_k(n)\} \quad \forall k \in S_T$
3. Compute $V(n) = \frac{1}{1+M(n)}$ for each draw n, where
$$M(n) = \sum_{k=1}^{4} \exp\left(\nu_k(n) \times \rho_k\right), \tag{B1}$$
   where ρ_k is the scaling vector.
4. The average of $V(n)$ over the N draws and over consumers and products gives the simulated likelihood function.

Similar to Ursu, Wang, and Chintagunta (2020) and Ursu, Zhang, and Honka (2023), we use different scaling values ρ_k for each of the decisions consumers make. Using our Monte Carlo simulation that closely resembles the estimation data, we determined that the following scaling parameters recover the data well: $\rho = [-0.69, -1.49, -0.39, -0.39]$ for discovery, search, stopping, and choice rules, respectively. Therefore, we estimate our model with the same set of scaling values.

As explained in Section 4 in the paper, before discovering a set of products, the consumer's expectations are given by $G_r(\cdot)$, which represents the normal distribution with mean μ_{δ_r} and standard deviation σ_{δ_r}. We obtain both parameters of the $G_r(\cdot)$ distribution from the data. More precisely, outside of the estimation, we compute the mean of each x_j variable that enters the model (price, a promotion dummy, an indicator for whether the consumer has searched products in our focal category in the previous week, and an indicator for whether the consumer has visited the sales page of the website in the previous week) for each route, multiply them by a parameter vector obtained from a preliminary estimation of the model, and calculate μ_{δ_r} by summing across the obtained values.[16] To compute σ_{δ_r} for each route, we calculate the standard deviation of the x_j's multiplied by the same parameter vector.

[16]Ideally, we would compute these values during the estimation for every draw of the parameter vector. However, the computation of these values affects the look-up table we need to compute discovery reservation utilities, and to compute this look-up table it takes longer than 3 minutes for each parameter draw. Therefore, to make our estimation feasible, we use the method described in this appendix.