Search Gaps *

Current version: December 2020

Raluca M. Ursu[†] Qianyun Zhang[‡] Elisabeth Honka[§]

Abstract

In the canonical sequential search model, consumers inspect options consecutively until they decide to stop searching, a decision which occurs only once before consumers determine whether and what to purchase. However, using data on consumers' online browsing histories, we document that consumers frequently take breaks during their search ("search gaps"), that is, they obtain information on a number of options, pause, and later resume their search. Further, we provide model-free evidence that consumers take breaks from searching due to fatigue. To describe search processes that include gaps due to fatigue, we extend the Weitzman (1979) framework and develop a sequential search model that rationalizes search gaps by allowing consumers to additionally decide when to search an option: now or after a break. Fatigue enters the model through increasing search costs: the more a consumer searches, the higher her search costs per option; taking a break reduces these costs to a baseline and enables the consumer to resume her search at a later time. We estimate the proposed model using our data and quantify the effect of fatigue on consumer search and purchase decisions. We find the effect of fatigue to be larger than that of baseline search costs. Lastly, using counterfactuals, we demonstrate the managerial importance of consumers' search fatigue.

Keywords: Sequential Search, Search Fatigue, Search Delay, Online Browsing, Fashion Industry

JEL Classification: D83, L81, M31

*We are thankful for comments from Tomomichi Amano, Simon Anderson, On Amir, Heski Bar-Issac, Alix Barasch, Bryan Bollinger, David Brown, Daria Dzyabura, Eric Greenleaf, Sarah Komisarow, Kanishka Misra, Eitan Muller, Brennan Platt, Anita Rao, Andrew Rhodes, Robbie Sanders, Daniel Savelle, Sandro Shelegia, Andrey Simonov, James Smith, Marton Varga, Ken Wilbur; seminar participants at Rotman School of Management (Toronto) and Rady School of Management (UCSD); and attendees of the 2019 Marketing Science/INFORMS conference and the 2020 Consumer Search Digital Seminar Series. We have greatly benefited from the careful and meticulous help offered by Ishita Verma, Jian Zhang, and Xinyu Wei in augmenting and processing our data. We would also like to thank Siham El Kihal and GfK for helping us obtain the data. We appreciate the funding received from the Center for Global Economy and Business, NYU Stern. The usual disclaimer applies.

[†]Stern School of Business, New York University, rursu@stern.nyu.edu.

[‡]Stern School of Business, New York University, qianyunzhang@stern.nyu.edu.

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

1 Introduction

Consumers typically face many options before making a purchase decision. For example, consumers can choose among 1,000 TVs under $500 on Amazon, more than 400 hotels in New York City on any given night, and almost 45 new car models every year.[1] Since evaluating each product is costly, consumers typically only investigate a small number of options. Models of consumer search describe how individuals decide which options to become informed about and which ones to ignore (Stigler 1961, Weitzman 1979). These models characterize consumer search behavior as an *uninterrupted* process, with one stopping decision occurring after the consumer has evaluated all options that are optimal to be inspected.

However, using data on consumers' entire online browsing histories, we document that search processes frequently involve breaks ("search gaps"). That is, consumers often obtain information on a number of options during a session, then take a break from searching, and later resume their search in a different session (e.g., a day later). Our data come from GfK and capture all web traffic (8 million clicks) of a panel of 4,600 Dutch consumers during ten weeks in 2018. The data include all clicks in our focal category – fashion – as well as all other browsing activities that consumers performed during the same session (e.g., checking emails, visiting Facebook, or using search engines). Importantly, consumers' online activities can be linked over time, i.e., across sessions, revealing when consumers search products versus when they take a break from searching. With these data, we show that search gaps are prevalent: on average, 43% of consumers take at least one break while searching. Further, conditional on pausing the search at least once, the average search process contains three gaps.

Such search gaps are ignored by previous consumer search literature for several reasons. First, prior empirical work often does not observe search gaps because it employs data containing information on purchases only, survey data with information on purchases and only searched sets, or browsing data at the session level that cannot be linked across sessions or to an individual consumer(e.g., Seiler 2013; Honka 2014; Honka and Chintagunta 2017; Ursu 2018). And second, even when data on search gaps are available, prior literature assumes that sessions are either independent or that they can be grouped together as part of a (gap-free) larger search (e.g., Chen and Yao 2017, De los Santos and Koulayev 2017; Ursu, Wang, and Chintagunta 2020).[2]

[1] The first two statistics are obtained from Amazon.com and Expedia.com, respectively, while the last one is obtained from https://www.statista.com/statistics/200092/total-number-of-car-models-on-the-us-market-since-1990/.

[2] Another way of interpreting the decision to ignore search gaps is to say that prior work has assumed that only the last stopping

Assuming away search gaps is innocuous if the reason for such gaps is unrelated to consumer search decisions, e.g., work emails or planned offline activities. However, there are several reasons to expect that search gaps are – to a large extent – deterministic. For example, search gaps may occur when consumers expect prices or product features to change over time and thus think that they may benefit from delaying their search. A second potential reason is shopping fatigue: the more options the consumer searches, the higher her search costs per option due to fatigue; taking a break reduces these costs and enables the consumer to resume her search at a later time.[3]

Using model-free evidence, we show that search gaps are largely deterministic and that fatigue is a main driver of search gaps in our data. To demonstrate the first point, we investigate what consumers do during search gaps. We find that approximately 90% of consumers engage in other online activities after ceasing their fashion search during a session. Such activities are often leisure activities (e.g., visiting social networking websites), suggesting that consumers had more uncommitted time during which they could have continued searching fashion products, but that they chose not to do so. Also, we find that consumers often resume their fashion search within a session when interrupted by an email notification, further supporting the idea that search gaps are a consumer's choice.

Next, we provide three pieces of evidence in support of the notion that search gaps are related to fatigue. First, we proxy for fatigue using consumer demographics, time variables, and website characteristics. We show that consumers who are older, who search predominantly during the day, and who visit websites that are slower to load or are harder to read have generally more search gaps. Second, we show that the more websites a consumer has searched and the more time she has spent searching since the last break, the higher her likelihood of a gap. Finally, we provide empirical evidence against several alternative explanations for the occurrence of search gaps: (i) expecting future changes in prices or other product features, (ii) having a limited budget of time, or (iii) forgetting previously obtained information.

We then develop a model of sequential search that endogenizes search gaps. More specifically, we extend the Weitzman (1979) sequential search model in two directions. First, we allow consumers to not only decide which products to search and in what order, but also *when* to search them: now or after a break. Second, motivated by our empirical evidence, we allow the decision of when to search a

decision is relevant.

[3]The following quote illustrates shopping fatigue:"Car shopping is exhausting and confusing. With every search online, I have to drink a sip of wine." (http://business.time.com/consumer-fatigue-shopping-has-never-been-easier-or-as-mentally-exhausting/).

product to be influenced by fatigue. To this end, we model search costs as having two components: a baseline level and a component that depends on the number of options searched after the latest break. This means that the cost of gathering product information after taking a break is equal to the baseline search cost level, while subsequent searches involve paying a higher search cost per option due to the fatigue that accumulated from previous searches within the same session. Note that these two components affect consumer decisions differently. For example, higher fatigue levels increase the number of gaps, while higher baseline search costs (indirectly) decrease the number of gaps by reducing the number of searched options.

The model we develop captures search gaps, but – in contrast to the Weitzman (1979) problem – no longer has an index policy solution. This is the case because, in our model, the optimization problems of different options interact for two reasons: (i) due to the increasing nature of search costs, i.e., searching an alternative increases search costs for *all* so far unsearched alternatives, and (ii) due to the choice of when to search an option, i.e., choosing to take a break resets search costs for *all* unsearched options. This interaction of optimization problems of different options violates the assumption in Weitzman (1979) that searching an option does not affect the payoffs from any other option, leading to a failure of the index policy solution. However, we show that, if the value of continued search satisfies a monotonicity condition, then the optimal search order in our problem coincides with the one in Weitzman (1979). Using this result, we then describe a consumer's optimal search rules for the entire set of decisions she makes in a model with search gaps: (i) which alternatives to search, (ii) when to search an alternative, and (iii) whether to continue searching or to stop. These optimal search rules are characterized by a set of three reservation utilities (rather than one reservation utility as in Weitzman 1979) – one for each of the three decisions a consumer makes.

We estimate our model and quantify consumer preferences, baseline search cost, and search fatigue in two fashion subcategories, "shirts, tops, & blouses" and "shoes." Our empirical results are consistent across both fashion subcategories. We recover consumers' utilities for the 10 most popular websites and find that consumers are loyal to websites they have visited frequently before. More importantly, fatigue has a large effect on search decisions, equivalent to increasing baseline search cost at least tenfold with every searched option. In contrast, the baseline search cost estimate is relatively small. However, estimating the model using the canonical Weitzman (1979) framework leads to an overestimate of the baseline search cost. The Weitzman model ignores search gaps and assumes search

costs are independent of the number of previously searched options. Thus, when estimated on the same data as a model where fatigue affects search costs, the Weitzman model rationalizes the same number of searched products by inflating baseline search costs. We also show that an adapted version of the Weitzman model that ignores search gaps, but models fatigue, i.e., search costs are an increasing function of the number of previously searched options, also overestimates baseline search costs, albeit less, and underestimates the effect of fatigue. This occurs because, ignoring the fact that fatigue caused consumers to take a break from searching (not only increased search costs), leads to the mistaken impression that fatigue has a relatively smaller impact on consumer decisions.

Finally, via counterfactuals, we address two managerially relevant issues related to search fatigue. First, we study how much companies benefit from a reduction in search fatigue and whether these benefits are heterogenous across companies. We find that decreasing fatigue by 50% increases the number of searches by $1-4\%$, increases transactions by $0.5-1.2\%$, and lowers search gaps by at least 11%. Interestingly, this effect is larger than that of decreasing baseline search costs by 50%, a policy that most prior work focuses on (e.g., Seiler 2013, Honka 2014). Further, we observe that smaller and less popular websites benefit more from a reduction in fatigue than larger and more popular websites. This is the case because a reduction in fatigue leads to additional searches and purchases that would not have occurred on smaller websites if fatigue levels were higher. And second, we investigate the consequences of consumers not being able to reduce their fatigue levels via a search gap. Such a situation might occur when consumers face challenging times such as the Covid-19 pandemic or when consumers are constantly being stimulated by (tiring) marketing activities. We find that not being able to reset fatigue during a break leads to a significant reduction in the number of products consumers search and purchase: searches decrease by approximately 20% and purchases by more than 6%. Most importantly, larger and more popular websites are hurt less in such a situation. In other words, consumers become more likely to buy from larger and more familiar websites. This finding emphasizes the importance of search gaps for competition and brand value.

Our paper makes several contributions to marketing research and managerial practice. In terms of theory, we document the presence of search gaps during consumers' search processes and propose the first model of consumer search that accounts for such search gaps. To the best of our knowledge, we are also the first to model consumer search fatigue before a purchase.[4] From a managerial perspective,

[4]Carlin and Ederer (2019) develop a model of search fatigue in which fatigue affects consumers across purchases rather than before a purchase, which is our focus.

our approach can help companies measure search fatigue, which is the first step in reducing it. Lower fatigue levels encourage consumers to search and purchase more products. Furthermore, search gaps reveal that consumers might be stopping their search due to a high fatigue level rather than a low match value with a brand. Identifying and targeting such consumers might be profitable since these consumers are still active in the market and thus more likely to resume searching and ultimately purchase (Schmittlein, Morrison, and Colombo 1987). The observation that consumers may stop searching because of a high fatigue level also has implications for firms' pricing decisions. Prior work on ordered search has found that firms' optimal pricing decisions depend on the order in which they are searched (e.g., Arbatskaya 2007; Armstrong, Vickers, and Zhou 2009; Petrikaite 2018). However, these results no longer necessarily hold when consumers can take search breaks.[5]

The rest of the paper is organized as follows. In the next section, we discuss relevant prior work. In Section 3, we introduce our data and in the following section, we provide model-free evidence for search gaps being deterministic and related to fatigue. We develop our theoretical model in Section 5. In Section 6, we describe our empirical model, estimation procedure, and identification. In the following section, we present our results, while, in Section 8, we provide managerial implications using two counterfactual exercises. We conclude in the last section.

2 Relevant Literature

This paper is primarily related to three strands of the literature: (i) the theoretical consumer search literature, (ii) empirical work using individual-level search data to quantify consumer preferences and search costs, and (iii) prior work on choice deferral. We describe and delineate our paper vis-à-vis extant research.

We contribute to theoretical work on consumer search in two main ways. First, we develop a new model of consumer search, adding to a rich literature that generally follows one of two frameworks: either the sequential search model of Weitzman (1979) or the simultaneous search model by Stigler (1961). In both these frameworks, consumers inspect products consecutively until they decide to stop

[5]For example, Armstrong, Vickers, and Zhou (2009) show that the non-prominent firm can infer that the consumer searching it obtained a low match value at the prominent firm. In this case, the non-prominent firm will face a relatively more inelastic demand for its product, allowing it to charge a higher price than the prominent firm in equilibrium. This inference is weakened when the consumer has the option to visit the non-prominent firm after a search gap, since such a decision may be motivated by the low search cost after the break, rather than a low match with the prominent firm. Thus, observing when the consumer searches an option (before or after a search gap), not only whether she searches, may help companies' pricing strategies.

searching, a decision which occurs once before determining whether to purchase. For example, in Weitzman (1979)'s sequential search model, the consumer proceeds to searching options as long as the benefit from searching exceeds the cost. When this relation no longer holds, search ceases and the consumer determines whether to purchase. Similarly, in Stigler (1961)'s simultaneous search model, there is one stopping decision: after searching the set of options for which the expected benefit exceeds search cost, the consumer stops and decides whether to buy one of the searched products. Thus – in contrast to our model – neither framework can be used to study search gaps, which involve multiple stopping decisions.

The only exception is a model for homogenous goods developed by Morgan and Manning (1985). The authors demonstrate that, under very general conditions, neither simultaneous nor sequential search is optimal, but rather a combination of the two is, i.e., a process during which the consumer searches sets of options sequentially. Morgan and Manning (1985)'s model can give rise to search gaps since consumers may choose sets of options to search at every occasion and take breaks between sets. However, since their theory was developed for homogenous goods, i.e., all products are ex ante identical, it can explain how consumers choose the number of options to search in every set, but not the identity of those options. Therefore, to the best of our knowledge, no prior theoretical work exists that can account for search gaps when consumers search among heterogenous goods, i.e., also choose which products to search. Our model fills this gap in the literature.

Second, we contribute to the theoretical consumer search literature through our definition of search costs. Most prior work assumes that search cost per product are independent of the number of products searched.[6] To the best of our knowledge, there are only a few exceptions. Stiglitz (1987) studies the effect of convex search costs on competition and the equilibrium number of firms in the market. The author links this effect to the increasing scarcity of time and money that intensifies as the consumer continues searching, but does not consider the effect of resetting these costs on search decisions. Levav et al. (2010) show experimentally that participants who need to customize a product (a suit or a car) are more likely to choose the default option when first presented with options that have many rather than few attributes. The authors argue that this result can be partially explained by convex costs of evaluating attributes, as demonstrated by literature in psychology and economics modeling self-control as a muscle that requires more effort on future rather than identical early stimulation

[6] For a review of theoretical work on consumer search, see Baye, Morgan, and Scholten (2006) and Anderson and Renault (2018).

(Ozdenoren, Salant, and Silverman 2012; Vohs et al. 2008).

Carlin and Ederer (2019) develop a model of search fatigue in which fatigue affects search decisions across purchase trips, i.e., the more products the consumer searched before the previous purchase, the higher her search costs are when searching towards the next purchase decision. In contrast, in our paper, we focus on the effect of fatigue on search decisions before a given purchase, i.e., the more the consumer searches before the current purchase, the higher her search costs.[7] Also, in Carlin and Ederer (2019) the goal is to study the effect of search fatigue on firm pricing decisions in equilibrium, while we take our model to data and quantify the effect of fatigue on consumer decisions. Most closely related to our paper, Ursu and Dzyabura (2020) posit that search costs increase linearly in the number of alternatives searched and affect current search and purchase decisions, modeling choices which we also make. However, the presence of increasing search costs is not sufficient for the occurrence of search gaps. More precisely, such search costs may explain why the consumer stops searching, but not why she restarts. For consumers to be willing to resume their search, search costs must also decrease during a gap (if the consumer's utility from the available options remains unchanged). To the best of our knowledge, no prior work on consumer search suggests this possibility. Instead, prior economics work on education finds that taking a break from academic classes to perform physical exercises, helps students to recover from cognitive fatigue and to perform better academically (Bednar and Rouse 2019). We posit that a similar mechanism may drive search fatigue.

Our paper is also related to empirical work quantifying preference and search cost parameters using individual-level data on consumers' search activities (e.g., De los Santos, Hortaçsu, and Wildenbeest 2012; Honka 2014; Koulayev 2014; Chen and Yao 2017;Honka and Chintagunta 2017; Honka, Hortaçsu, and Vitorino 2017; De los Santos and Koulayev 2017; Ursu 2018). [8] Most of this work assumes that search costs per product are independent of the number of products searched. The exception is Koulayev (2014) who estimates higher search costs for products searched later, providing empirical support for the assumption of increasing search costs. Furthermore, this stream of literature rests on the theoretical models of Weitzman (1979) and Stigler (1961) and assumes that consumers search options consecutively and stop searching only once. Although some prior work recognizes the fact that consumers search in sessions (e.g., consumers learn across sessions in Wu et al. 2015), it does not

[7] A concept that may seem similar to fatigue is that of obfuscation (e.g., Ellison and Ellison 2009, Ellison and Wolitzky 2012). However, the difference is that by obfuscating the consumer, firms increase their (baseline) search costs rather than making successive searches more costly. As such, our model and predicted behavior differs from those observed in a model with search obfuscation.

[8] For a review of empirical work on consumer search, see Honka, Hortascu, and Wildenbeest (2019).

explicitly model consumers' decisions to stop and resume searching several times, and is thus not accounting for the presence of search gaps.

Finally, our paper relates to the literature on choice deferral. Work in consumer behavior shows that choice difficulty increases the probability of the consumer choosing none of the options and thus delaying her choice (Dhar 1997; Novemsky et al. 2007). In the context of a search model, we view this finding as broadly suggesting that search gaps are more likely as search difficulty increases, a result which is in line with our empirical patterns. More closely related is the work of Greenleaf and Lehmann (1995) that identifies several possible reasons for consumers delaying the decision to purchase a product such as the absence of time to devote to the task or the expectation of future price decreases. Although not described in the context of consumer search, these reasons could also influence search decisions. We contribute to this literature by developing a model of consumer search in which consumers may stop and restart searching, thereby formalizing the idea of delay in the context of search.

3 Data

3.1 Data Sources

Our primary data come from GfK, Germany's largest market research company. GfK recruits and maintains an online panel of representative consumers for whom online browsing data are collected via a browser extension. This browser extension is installed on the panelists' devices (PC, smartphone, tablet) and records all their online activities. GfK groups all clicks which are not interrupted by a time period of inactivity longer than 30 minutes (the industry standard) into "sessions." The data are at the exact URL level clicked by a consumer and also contain the time of each click, the visited website, and consumer demographics (e.g., age, gender). Furthermore, GfK classifies clicks into activities such as email, social networking, fashion, search engine use, banking, or gaming. And finally, GfK codes the transaction funnel identifying website visits, product views, basket additions, checkouts, and order confirmations.

Our data contain the complete PC browsing histories of online panel members from the Netherlands from February 15, 2018, until May 1, 2018, for sessions during which they made at least one click to a fashion website. In other words, our data are conditional on a fashion click (not conditional on a

purchase) occurring during a session, but show all visited websites (including non-fashion websites) during such sessions. We chose to focus on products in the fashion category for two reasons: first, this category is frequently visited by consumers, allowing us to observe multiple search actions. And second, we were able to scrape product information for the URLs in the GfK data because they are stable over time and are generally not personalized to individual consumers.[9] While choosing a category such as travel would allow us to observe enough search activity, we would not be able to scrape product information since this information changes frequently and is often personalized. On the other hand, a durable goods category only contains searches from a small number of consumers and would restrict our analysis given our relatively short observation window.

We augmented the GfK data in several ways. First, we scraped product information from 44 of the top 50 fashion websites. These 44 websites account for more than 57% of all fashion clicks, a large percentage given the 1,046 unique fashion websites in our data. This data collection stage occurred within one month of the last day of our observation period to prevent changes on the webpages. The product information we obtained includes price (current and any promotions), page title, brand name, product name, product color, reviews, star rating, number of photos, product description, shipping information, speed score of the website, word counts, and page readability.[10] Second, we identified the purchased product as the last product searched before engaging in transaction related clicks (e.g., adding to cart, checking out, confirming an order) on the same website. Using this information, we defined a "spell" as all search sessions conducted by a consumer before a purchase (or before the end of our observation period if no purchase occurred). Next, we use URLs, page titles, and the scraped information (e.g., product description) to identify nine product subcategories (e.g., "shoes" or "accessories") that the consumer searched. Finally, we defined a "search gap" as the break a consumer takes between subsequent search sessions. Figure 1 provides an example of a search process, defining the concepts we use in this paper. Detailed information about the data collection, classification, and cleaning steps are provided in Web Appendix A.

[9] Cavallo (2017) finds that 92% of fashion prices are the same online and offline within a chain, suggesting little personalization to individual consumers' visits.

[10] We obtain website speed score information from Google https://developers.google.com. The website speed score is the page loading speed with values ranging from 0 to 100. Google PageSpeed Insights considers $0-49$ as slow, $50-89$ as medium, and greater than 90 as fast speed. We obtained other website features such as word counts, number of images, and readability from `https://urlprofiler.com/`. Page readability is measured in terms of its SMOG index, which computes the number of years of education needed to understand a piece of text. Therefore, a larger SMOG index means a less readable text. More information about the SMOG index can be found at `https://en.wikipedia.org/wiki/SMOG`.

==========

Insert Figure 1 about here

==========

3.2 Data Description

Our data contain 7,877,551 observations with 428,651 fashion clicks. We observe searches made by 4,622 consumers in 5,665 spells and 40,735 sessions across nine distinct fashion subcategories. There are a total of 3,036 products purchased in the fashion category, with 76% of spells containing no purchased product, 11% of spells containing one purchased product, and 13% of spells containing at least two purchased products. 65% of consumers are female; the average (median) age is 48 (49) with a large standard deviation of 16. Click duration is, on average, half a minute and is slightly longer for fashion than non-fashion clicks (0.54 versus 0.50 minutes, respectively).

We summarize session and spell characteristics in Table 1. Activity in each session is extensive: on average, consumers make 190 clicks on 30 websites and spend more than one hour online. In contrast, fashion search in a session is more modest: the average consumer makes 11 fashion clicks, spends about five minutes searching, and visits one fashion subcategory. The most popular activities in our data are email, social networking, and fashion. Together they account for more than 33% of all clicks. The most popular websites are google.com, live.com, and facebook.com.

==========

Insert Table 1 about here

==========

Table 2 lists the top fashion websites in terms of their searches and transactions. Zalando is the most popular fashion website in our data and among online retailers in the Netherlands.[11] More precisely, Zalando has more than 22% of transactions in our data (15% of all fashion clicks), followed by H&M with 13% of transactions (10% of all fashion clicks). The two most commonly purchased fashion subcategories are "shirts, tops, & blouses" and "shoes." In Table 3, we additionally display the most popular websites searched and purchased in each of these two subcategories. Once again we note the overall popularity of Zalando as well as C&A and H&M in the "shirts, tops, & blouses" subcategory, and of Schuurman Shoenen and Van Haren in the "shoes" subcategory. Finally, the

[11] For details, see https://ecommercenews.eu/top-10-online-stores-in-the-netherlands/.

subcategory "jackets & vests" is the most expensive one with an average transaction price of 60€, while "children's clothes," the cheapest subcategory, has an average transaction price of less than 20€.

==========

Insert Tables 2 and 3 about here

==========

In Table 4, we demonstrate that search gaps are very prevalent in our data. More specifically, across all nine fashion subcategories, on average, 43% of search spells contain at least one search gap. Conditional on a spell having at least one search gap, the average number of search gaps is 3 per spell. For this reason, our paper focuses on studying why search gaps occur and how they can be understood from the lens of a search model.[12] The average length of a search gap (number of days between search sessions) is approximately one week. Spells range from 7 to 17 days, on average, depending on the subcategory.[13] In comparison, the average time between spells (for the approximately 30% of consumers who have more than one spell during our observation period) is longer, typically lasting about two weeks.

==========

Insert Table 4 about here

==========

To summarize, we find that search gaps occur frequently both across consumers and fashion subcategories. In the next section, we delve deeper and attempt to relate search gaps to consumer fatigue.

4 Empirical Evidence Relating Search Gaps to Fatigue

In this section, we first provide evidence that search gaps are deterministic, i.e., that consumers *choose* when to search each option. Second, we show empirically that these decisions are affected by fatigue. And finally, we provide evidence against possible alternative explanations for the occurrence of search gaps.

[12] We also observe when a consumer takes a break within a session. 31% of sessions contain such gaps which, on average (median), only last 3.73 (0.27) minutes. Because such gaps are less prevalent and are unlikely to lead to a change in fatigue due to their short duration, our focus in this paper is on analyzing search gaps that happen across sessions.

[13] The average spell length may be lower than the average search gap length because only a fraction of spells contain gaps.

4.1 Empirical Evidence Showing Search Gaps are Deterministic

Although we do not directly observe the consumer's decision process, there are several pieces of evidence in our data pointing to search gaps being deterministic, i.e., occurring as a result of a decision by the consumer to delay her search. First, we find that consumers' online activity rarely ends when their search in the fashion category ends. More specifically, only 13% of sessions end with a fashion click. Furthermore, the two most popular activities after the last fashion click are email and social networking, accounting for more than 23% of all clicks. These two categories remain the most popular even when restricting to clicks in the evening (6pm to midnight) or on the weekend, increasing the chances of them capturing leisure activities. This suggests that the consumer had more time available to allocate to online activities, but chose not to spend more time searching in the fashion category and that search gaps are a result of a conscious choice made by the consumer.

In addition, although we do not observe what consumers do during a search gap (across sessions), we observe what they do when pausing their search within a session. Here again, we find that email is the most popular activity, with 16% of clicks, followed by social networking, with approximately 6% of clicks.

Finally, we observe when a notification announcing that an email was received interrupts search and how consumers react to this event. Receiving such a notification is arguably exogenous to the consumer search for fashion products. We find that 91% of consumers, who get a notification after searching in the fashion category, return to searching fashion in the same session. In other words, consumers do not pause their search in response to an email notification, i.e., take a search gap, but rather choose when to take a search break.

4.2 Empirical Evidence for Fatigue Affecting Search Gaps

Here, we aim to link fatigue and search gaps. Doing so is difficult because we do not directly observe consumer fatigue levels, so we cannot observe whether they relate to the decision to take a break versus to continue searching without a break. An ideal experiment would manipulate consumers' fatigue levels (e.g., by making some websites slower to load or by increasing the amount of information available) and directly test whether fatigue affects search gaps. In what follows, we seek to mimic this experiment and test the relation between fatigue and search gaps using observational data.

First, since we do not directly observe the fatigue level of a consumer, we consider three proxies

for it: (i) consumer demographics, (ii) time variables, and (iii) website characteristics. We then check whether these fatigue proxies are related to the number of gaps a consumer makes while searching. We define our dependent variable as the logarithm of the number of search gaps in a spell (plus one). In Table 5, we present our results. The first two columns show the results for "shirts, tops, and blouses" and the last two columns show the results for "shoes." All regressions control for the number of searches consumers performed, since consumers who search longer generally also have more search gaps.[14]

==========

Insert Table 5 about here

==========

We find that older consumers have more search gaps in a spell. There is an abundance of medical research supporting the idea that mental processing abilities are affected by age, with observed declines in conceptual reasoning, memory, processing speed, and attention to stimuli in older individuals (Harada, Love, and Triebel 2013). These changes in mental abilities can affect decision-making processes in marketing-relevant contexts (Peters 2010; Carpenter and Yoon 2011). For example, older consumers have been shown to make better decisions when presented with fewer options (Abaluck and Gruber 2013; Tanius et al. 2009). Also, research shows that older consumers are more likely to use heuristics, to search for a shorter amount of time, and to build smaller consideration sets to reduce cognitive effort (Lambert-Pandraud, Laurent, and Lapersonne 2005; Kim et al. 2005). Motivated by this evidence, we consider age as a possible proxy of a consumer's proneness to fatigue and find age to have a positive effect on search gaps. We find mixed evidence for the effect of gender on search gaps: for "shirts, tops, and blouses," we find women to take significantly more breaks than men, but we find the effect of gender to be insignificant for "shoes." Moreover, we find that consumers who predominantly search in the evening (6pm to midnight), when there are likely fewer constraints on their time, have fewer search gaps, with mostly insignificant results for weekend searchers. Finally, consumers who visit websites that are slower to load (lower speed score) and harder to read (higher readability/SMOG index) have more search gaps, while the number of images and words on the page have little to no effect.[15] Assuming that these measures are suitable proxies for consumer fatigue, our results show that higher fatigue levels lead to more search gaps.

[14] Results without controlling for the number of searches consumers perform are largely unchanged and available from the authors upon request.

[15] The correlations between these website characteristics are smaller than 0.2.

Second, we check whether the number of options searched after the last gap affects the probability of a gap. If more search increases fatigue levels, then the more websites the consumer searches after the latest break, the higher the probability of a search gap. We consider two measures of the number of searches performed: (i) the cumulative number of websites searched since the previous gap and (ii) the total number of minutes spent searching since the previous gap. Our results are displayed in Table 6. The analysis controls for spell and website fixed effects as well as for sessions that end in a transaction. Consistent with our hypothesized relation between fatigue and search gaps, we find that the more websites a consumer searched after the latest break or the more time she spent searching, the more likely it is for the consumer to take a break.

==========

Insert Table 6 about here

==========

4.3 Empirical Evidence Against Alternative Mechanisms

In what follows, we briefly discuss several potential alternative explanations for the occurrence of search gaps. Our goal is not to rule out *all* alternative explanations. Indeed, multiple reasons may affect the occurrence of search gaps. Rather, our goal is to provide empirical evidence against some of the most common alternatives. Modeling the relation between search gaps and factors other than fatigue is left for future research.

One alternative explanation for search gaps is that consumers may delay their search because they expect prices to decrease or other product features to improve. This reason is particularly pertinent in a category such as travel, where airfare and hotel prices change very frequently and dynamically in response to changes in demand and the available supply of options. However, in the fashion category, price and product feature changes are less frequent. More precisely, product features change mostly every season and there are only two major seasons annually (Fall/Winter, running from July to December, and Spring/Summer, running from January to June),[16] while prices typically change around holidays or at the end of each season when products go on sale.[17] Our observation period (February

[16] For more details, see https://www.leaf.tv/articles/when-do-fashion-seasons-start/.

[17] Spring/summer goods usually goes on sale in June and July and Fall/Winter goods usually go on sale in January after the winter holidays (see https://money.usnews.com/shopping-holidays-the-best-days-to-shop-this-year, https://www.thebalance.com/comprehensive-guide-to-seasonal-sales, or for the Netherlands https://www.amsterdamsights.com/shopping/sales-period).

15 to May 1) falls within a single season and does not overlap with any major sales periods.[18] Also, the average search gap in our data is approximately one week long. Therefore, in our empirical setting, it is unlikely that search gaps occur because consumers expect prices or other product features to change, since such changes are typically separated by several weeks or months.

A second alternative explanation is that consumers resume their search to obtain additional information about the same products or because they forgot the information they searched previously. Note that these behaviors occur even absent gaps, i.e., they may occur even if the consumer does not take any breaks during her search. More importantly, both these decisions involve revisits of previously searched options (Ursu, Wang, and Chintagunta 2020; Dang, Ursu, and Chintagunta 2020). However, in our data, we find that product page revisits account for only 5.3% of all clicks and that most of these revisits (71.7%) occur within a session rather than across sessions. Therefore product page revisits cannot explain search gap behavior.

Finally, search gaps may occur because consumers only have a predetermined budget of time available to allocate to searching during the current session.[19] However, as previously noted, we find that only 13% of sessions end with a fashion click and that fashion searches are typically followed by leisure activities. This observation suggests that consumers had more uncommitted time available, but that they chose to not devote it to additional fashion searches.

In sum, in this section we showed that search gaps are largely deterministic and that they are affected by fatigue. Next, we develop a model of sequential search that endogenizes search gaps due to fatigue.

5 Model

5.1 Setup

A consumer seeks to purchase an option $j = 1, \ldots, J$ or to choose the outside option of not purchasing (denoted by $j = 0$).[20] The consumer knows the utility distributions $F_j(\cdot)$, but has to search to learn the actual utility u_j, i.e., u_j is an independent draw from the continuous distribution function $F_j(\cdot)$. Search

[18] The exception is Easter, which occurred on April 1, 2018. However, comparing transactions one week before and one week after Easter shows no significant difference in prices ($t = 1.32$, difference= 0.02), consistent with the idea that prices did not change during this period.

[19] Note that this reason differs from fatigue because of the predetermined nature of a budget of time.

[20] We omit consumer i subscripts in what follows, but our model should be understood to apply to every consumer.

occurs sequentially. Let S denote the set of searched options, while $\bar{S}$ denotes the set of options still available for search. The best option among the searched ones is denoted by y, i.e., $y = \max_{j \in S \cup \{0\}} u_j$. Consumer choices depend on the state variables $\bar{S}$ and y.

We extend the above framework developed by Weitzman (1979) to account for search gaps. This involves making two modifications. First, we allow the consumer not only to decide which option to search, but also when to search it: now or after taking a break. To this end, we define a new (additional) state variable t which tracks the number of options searched after the last break, implying $|S| \geq t$. If the consumer decides to continue searching, she can either search option j immediately with $t > 0$ or after taking a break, which resets t to zero.[21]

And second, we allow search cost per option to increase with the number of searched options.[22] Following Ursu and Dzyabura (2020), search cost are given by

$$c_j = c_{j0} + \alpha t \tag{1}$$

where c_{j0} are baseline search costs and t captures the number of searches without a break. Recall that $t = 0$ if the consumer takes a break. The first component c_{j0} captures the cost of searching j regardless of the number of other options searched. It depends on characteristics of that option, for example in the case of a website its prominence or design (e.g., user friendliness). The second component depends on t and represents the consumer's fatigue from searching, i.e. that due to the number of previously searched options. For simplicity, we assume that the difference in cost between subsequent searches is constant, i.e., it costs an additional $\alpha > 0$ for the consumer to search j after having searched t other options. This functional form implies that the cost of the first search and that of searching an option after taking a break equal c_{j0}, while other searches involve paying a higher cost per option.

Given the state variables $(\bar{S}, t, y)$, at every search occasion when $t > 0$, the problem solved by the consumer is given by

$$V(\bar{S}, t, y) = \max_{stop,\, continue} \left\{ y,\ \max_{j \in \bar{S}} \Phi_j(\bar{S}, t, y) \right\}. \tag{2}$$

[21] Many factors (e.g., advertising) may influence a consumer's decision of when to resume her search. However, modeling consumers' choice of the length of the search gap is beyond the scope of this paper and we assume we observe the consumer searching again after t was reset. Consistent with our modeling choices, a preliminary analysis shows that having a longer gap does not correlate with the number of searches performed in the next session. This suggests that only a short amount of time is needed to reset search costs, making it not paramount to model the gap length. Analysis available upon request.

[22] An alternative model would be to let search costs be a function of the elapsed time since the prior break. We leave such a model to future research. Instead and consistent with most of the literature that considers increasing search costs (e.g., Stiglitz 1987, Carlin and Ederer 2019, Ursu and Dzyabura 2020), we let search costs be a function of the number of searched options, and note the similarity between these approaches (i.e., greater elapsed time is a direct consequence of additional searches).

The search value $\Phi_j(\bar{S},t,y)$ is defined as

$$\Phi_j(\bar{S},t,y) = \max_{now,\, later} \left\{ -c_{j0} - \alpha t + W_j(\bar{S},t+1,y),\ \beta\left[-c_{j0} + W_j(\bar{S},1,y)\right] \right\}, \tag{3}$$

with $0 < \beta < 1$ being the discount factor if the consumer decides to search later. Furthermore, the continuation value $W_j(\bar{S},t,y)$ is given by

$$W_j(\bar{S},t,y) = V(\bar{S} \setminus j,t,y)F_j(y) + \int_y^{\infty} V(\bar{S} \setminus j,t,u)dF_j(u). \tag{4}$$

The interpretation of the value function in Equation (2) is as follows: given a set of options available for search $\bar{S}$, a number t of options searched after the latest break, and a best option observed so far y, the consumer makes three decisions. First, she decides whether to stop or to continue searching. If she stops searching, she gets a payoff y which represents the option of buying the alternative with the highest utility revealed among those searched or of choosing the outside option of not purchasing. Second, if she decides to continue searching, she has to make two decisions (simultaneously): (i) which option j to search among those not yet searched and (ii) whether to search the chosen option now (at t) or after a break. $\Phi_j(\bar{S},t,y)$ denotes the value of choosing to search j from $\bar{S}$. Searching j immediately involves paying a relatively high search cost due to $t > 0$. In contrast, choosing to search j after a break involves paying a lower search cost, but postponing search to a later time period, option which is discounted at a rate β. If the consumer decides to search j after taking a break, she receives no utility in the current time period and t resets to zero. In this model, a higher level of fatigue α encourages more search gaps, making the value of searching an option now less desirable. To ensure consistent behavior at every state, i.e., to ensure that a decision to postpone searching an option will be followed by the decision to actually search this option after a break (instead of abandoning search altogether), we focus on options for which $-c_{j0} + W_j(\bar{S},1,y) \geq 0$.[23]

5.2 Indexability and Related Problems

To the best of our knowledge, no index policy solution to the general problem presented in Equation (2) exists. To understand why this may be the case, consider the following examples of simpler versions

[23] Since the consumer chose to postpone searching j, it must be that $\beta\left[-c_{j0} + W_j(\bar{S},1,y)\right] \geq y$. This implies our condition $-c_{j0} + W_j(\bar{S},1,y) \geq 0$ since $y \geq 0$ and $\beta \geq 0$. Thus, after the break when $t = 0$, $-c_{j0} + W_j(\bar{S},1,y) > \max\left\{y, \beta\left[-c_{j0} + W_j(\bar{S},1,y)\right]\right\}$ proving that the consumer would actually want to search j after the break.

of our model: if gaps happen exogenously (i.e., the consumer does not choose whether to search now or later) and if search costs are constant, our problem coincides with the Weitzman (1979) problem, which means that an optimal index policy exists. The optimal search rules in this case are as follows: compute an index, i.e., a reservation utility, for each available option and proceed to search options in a decreasing order of these indices until all options have been searched or any unsearched options have an index lower than the best observed utility among those searched. However, if gaps happen exogenously but search cost increase with the number of searches, the problem above does *not* coincide with the Weitzman (1979) problem. In particular, it is not indexable because the optimization problems of different options interact: although utility draws are independent across options, searching an option increases the search costs for *all* so far unsearched options. This means that the assumption in Weitzman (1979) that searching an option does not affect the payoffs from any other option is violated, leading to a failure of the index policy solution. Therefore, one reason why no index policy solution exists in the general version of our problem is the increasing nature of search costs.

In addition, in our problem, consumers decide not only which options to search, but also when to search them (now or later), adding an additional layer of decision-making not present in the Weitzman (1979) model (or most other search models in the literature). This feature makes our model resemble a bandit superprocess (BSP), a generalization of a multi-armed bandit problem (MAB; itself a generalization of the Weitzman (1979) problem), in which a decision maker not only chooses which of a set of independent arms/processes to play in each time period, but also chooses from a number of actions for each arm/process conditional on playing (Gittins, Glazebrook, and Weber 2011).[24] Similarly, in our problem consumers choose not only which products to search, but also when to search them (now or later). BSPs are generally not indexable because the optimizations for different processes interact: choosing a certain action for one process can affect the rewards of a different process (Whittle 1980, Brown and Smith 2013). In our case, choosing to take a break resets search costs for *all* unsearched options, meaning that the payoffs from one option are affected by choices related to other options. Thus this additional layer of decision-making captured in our model also prevents indexability.

[24] Our model is not a bandit superprocess because the searchable products are not independent – they are connected through t.

5.3 Solution

Despite the aforementioned indexability issues, in this section, we propose a solution for the entire set of decisions a consumer makes in our model: (i) which alternatives to search, (ii) when to search an alternative, and (iii) whether to continue searching. The purchase decision remains the same (purchase the product with the largest realized utility among those searched) so we omit it in the following. We are able to derive an optimal solution after imposing an additional condition which we describe below.

Selection rule

The selection rule determines which option the consumer searches next if she decides to continue searching. As mentioned, if $t = 0$, search costs are constant and t does not affect the search process: the choice between searching an option now or later becomes deterministic, i.e., the consumer always searches the option now. In this case, our problem coincides with the Weitzman (1979) problem and is therefore indexable. Options are optimally searched in decreasing order of an index, i.e., a reservation utility. In contrast, if $t > 0$, the consumer may search options in a different order than the one in the Weitzman (1979) model as we show next.

Theorem 1. *The optimal search order when $t > 0$ may not coincide with the optimal search order when $t = 0$.*

Proof: To show this, it suffices to consider the search order of any two options, j and k. The value of searching j at t is given by $\Phi_j(\bar{S}, t, y)$ (see Equation (3)). Similarly, $\Phi_k(\bar{S}, t, y)$ denotes the value of searching k at t. For ease of exposition, we repeat Equation (3) for j and k in the following as well as provide some simplifying notation:

$$\begin{aligned}
\Phi_j(\bar{S}, t, y) &= \max_{now,\, later} \left\{ -c_{j0} - \alpha t + W_j(\bar{S}, t+1, y),\ \beta\left[-c_{j0} + W_j(\bar{S}, 1, y)\right] \right\} \\
&= \max_{now,\, later} \{a,\ \beta A\} \\
\Phi_k(\bar{S}, t, y) &= \max_{now,\, later} \{-c_{k0} - \alpha t + W_k(\bar{S}, t+1, y),\ \beta[-c_{k0} + W_k(\bar{S}, 1, y)]\} \\
&= \max_{now,\, later} \{b,\ \beta B\}
\end{aligned}$$

Suppose that, if $t = 0$, the consumer prefers searching j before k, i.e.,

$$\Phi_j(\bar{S}, 0, y) - \Phi_k(\bar{S}, 0, y) = W_j(\bar{S}, 1, y) - W_k(\bar{S}, 1, y) - \left(c_{j0} - c_{k0}\right) = A - B > 0. \tag{5}$$

We show that this does not necessarily imply that $\Phi_j(\bar{S},t,y)-\Phi_k(\bar{S},t,y)>0$ for any $t>0$, i.e., the search order when $t=0$ may be different from the search order when $t>0$. There are four cases to consider:[25]

1. Suppose $a>\beta A$ and $b>\beta B$. Then

   $$\Phi_j(\bar{S},t,y)-\Phi_k(\bar{S},t,y) \;=\; a-b \;=\; W_j(\bar{S},t+1,y)-W_k(\bar{S},t+1,y)-\left(c_{j0}-c_{k0}\right).$$

   This difference may be positive or negative.

2. Suppose $a>\beta A$ and $b<\beta B$. Then

   $$\Phi_j(\bar{S},t,y)-\Phi_k(\bar{S},t,y) \;=\; a-\beta B\,.$$

   Since $A>B$, it must be that $a>\beta B$ so the search order for any $t>0$ coincides with the search order when $t=0$.

3. Suppose $a<\beta A$ and $b>\beta B$. Then

   $$\Phi_j(\bar{S},t,y)-\Phi_k(\bar{S},t,y) \;=\; \beta A-b \;=\; \beta W_j(\bar{S},1,y)-W_k(\bar{S},t+1,y)-\beta c_{j0}+c_{k0}+\alpha t\,.$$

   This difference may be positive or negative.

4. Suppose $a<\beta A$ and $b<\beta B$. Then

   $$\Phi_j(\bar{S},t,y)-\Phi_k(\bar{S},t,y) \;=\; \beta A-\beta B\,.$$

   This difference is positive since $A>B$. Thus the search order for any $t>0$ coincides with the search order when $t=0$.

In sum, in cases 2 and 4, the same search order prevails for any t. However, for the other two cases, such a result is not generally true. □

From Theorem 1 we know that the selection rule in Weitzman (1979) may not describe the optimal order in which consumers search alternatives in our problem. However, the proof of this statement also reveals a condition under which the optimal search order in the two problems coincides.

Condition 1. *The difference in continuation values of two options j and k is said to be* ***monotonic*** *if, for any $t>0$, $W_j(\bar{S},t+1,y)-W_k(\bar{S},t+1,y) \;\geq\; W_j(\bar{S},1,y)-W_k(\bar{S},1,y)$ whenever searching j before k is optimal for $t=0$.*

In words, the condition requires that the difference between the continuation values of two products be monotonic in t, i.e., a constant or increasing function of t. If this is the case, then the consumer will not switch her optimal search order for different values of t leading to our next result.

[25] We only consider cases for which the expressions in Equations (5) hold with inequality since other cases are straightforward to solve based on these results.

Theorem 2. *Under Condition 1, the optimal search order when $t > 0$ coincides with the optimal search order when $t = 0$.*

Proof: Cases 2 and 4 in Theorem 1 do not require an additional condition to hold for the search order to be the same for any value of $t \geq 0$. In case 1, it is straightforward to see that Condition 1 is sufficient for the statement to be true. In case 3, under Condition 1, it follows that $a - b > 0$, which then implies $\beta A - b > 0$. Since this holds for any pair of alternatives, our statement follows. □

Note that $W_j(\bar{S}, 1, y)$ is the continuation value for $t = 0$. Therefore, it coincides with the continuation value in Weitzman (1979), $W_j(y) = yF_j(y) + \int_y^\infty u dF_j(u)$. This means that Theorem 2 describes when the search order in our model coincides with the search order in the Weitzman (1979) model. To paraphrase, Theorem 2 chronicles when consumers in our model search options in a decreasing order of reservation utilities as computed in Weitzman (1979), i.e., as the unique solution z_j to $c_{j0} = W_j\left(z_j\right) - z_j$. This fact allows us to state the selection rule for our problem as follows:

Selection rule. *Under Condition 1, if the consumer chooses to search an option at time t, then it will be the option $j \in \bar{S}$ with the highest reservation utility z_j, where z_j is the unique solution to $c_{j0} = W_j\left(z_j\right) - z_j$.*

<u>**Search rules**</u>

Under Condition 1, if the consumer decides to continue searching, she searches the option with the largest reservation utility, i.e., $j^* = \arg\max_{j \in \bar{S}} z_j$. Thus we can solve the problem in Equation (2) in two steps: first, the consumer determines j^*. And second, she solves Equation (2) for j^* and determines whether to stop searching, whether to search j^* immediately or whether to search it after a break. The search problem in the second stage reduces to

$$V(j^*, t, y) = \max_{stop,\, now,\, later} \left\{ y,\ -c_{j^*0} - \alpha t + W_{j^*}(\bar{S}, t+1, y),\ \beta\left[-c_{j^*0} + W_{j^*}(\bar{S}, 1, y)\right] \right\}. \tag{6}$$

Our next result on the optimal selection and search rules in our model follows directly from Equation (6) and Theorem 2.

Theorem 3. *Under Condition 1, the following selection and search rules are optimal:*[26]

[26] We break ties as follows: the consumer prefers to search now if choosing between any of the three options, and prefers to search later rather than to stop.

1. ***Selection rule:*** *order options in decreasing order of reservation utilities z_j (as defined by Weitzman 1979).*

2. ***Search rules:*** *if j is the option with the maximum reservation utility among the options not yet searched $\bar{S}$, then given (t, y)*

   - *search j now if* $-c_{j0} - \alpha t + W_j(\bar{S}, t+1, y) \geq \max\left\{y, \beta\left[-c_{j0} + W_j(\bar{S}, 1, y)\right]\right\}$;
   - *search j later if* $\beta\left[-c_{j0} + W_j(\bar{S}, 1, y)\right] \geq \max\left\{y, -c_{j0} - \alpha t + W_j(\bar{S}, t+1, y)\right\}$;
   - *stop searching if* $y \geq \max\left\{-c_{j0} - \alpha t + W_j(\bar{S}, t+1, y), \beta\left[-c_{j0} + W_j(\bar{S}, 1, y)\right]\right\}$.

Proof: As shown in Theorem 2, the selection rule above is optimal. The statements describing the search rules follow directly from Equation (6). Two other results remain to be shown.

First, it can be shown that if the consumer chooses not to search j, then she would also not search any other option with a lower reservation utility. Since the consumer would prefer to search j before k, it must be that $W_j(\bar{S}, 1, y) - W_k(\bar{S}, 1, y) - \left(c_{j0} - c_{k0}\right) > 0$. By Condition 1, $W_j(\bar{S}, t+1, y) - W_k(\bar{S}, t+1, y) - \left(c_{j0} - c_{k0}\right) > 0$ also holds. Together these imply that if $y \geq \max\left\{-c_{j0} - \alpha t + W_j(\bar{S}, t+1, y), \beta\left[-c_{j0} + W_j(\bar{S}, 1, y)\right]\right\}$, then $y \geq \max\{-c_{k0} - \alpha t + W_k(\bar{S}, t+1, y), \beta[-c_{k0} + W_k(\bar{S}, 1, y)]\}$, proving our statement.

Second, it can be shown that if the consumer chooses to search j later, she would not prefer to search another product k with a lower reservation utility now. More precisely, wanting to search j later requires that $W_j(\bar{S}, t+1, y) - \beta W_j(\bar{S}, 1, y) < (1-\beta)c_{j0} + \alpha t$, while wanting to search k now requires that $W_k(\bar{S}, t+1, y) - \beta W_k(\bar{S}, 1, y) \geq (1-\beta)c_{k0} + \alpha t$, which cannot hold true under Condition 1. □

5.4 Functional Form of Continuation Values

In the previous sections, we presented the most general version of the problem and its solution, proving optimal selection and search rules only using Condition 1. However, because the exact relation between $W_j(\bar{S}, t+1, y)$ and $W_j(\bar{S}, 1, y)$ is still largely unspecified, the model in its current format is difficult to bring to data. Therefore, in what follows, we make a simplifying assumption for the functional form of these continuation values. Under Condition 1, the difference in continuation values of two alternatives remains constant or increases as a function of t. Consistent with this condition, we assume that such differences are constant across products:[27]

[27] Consistent with the fact that $W_j(\bar{S}, t, y) \geq W_j(\bar{S}, 1, y)$ for any value of $t > 0$, Assumption 1 requires that $\delta \geq 0$. To see the former statement, recall that $W_j(\bar{S}, 1, y) = yF_j(y) + \int_y^\infty u dF_j(u)$. By definition, $W_j(\bar{S}, t, y) = V(\bar{S} \setminus j, t, y)F_j(y) + \int_y^\infty V(\bar{S} \setminus j, t, u) dF_j(u) = \max\left\{yF_j(y), \max_{k \in \bar{S} \setminus j} \Phi_k(\bar{S} \setminus j, t, y)F_j(y)\right\} + \max\left\{\int_y^\infty u dF_j(u), \int_y^\infty \max_{k \in \bar{S} \setminus j} \Phi_k(\bar{S} \setminus j, t, u) dF_j(u)\right\} \geq W_j(\bar{S}, 1, y)$, proving our statement.

Assumption 1. $W_j(\bar{S}, t+1, y) \;=\; W_j(\bar{S}, 1, y) \;+\; \delta t$ *for* $\delta \geq 0$ *and* $\forall j \in \{1, \ldots, J\}$.

This assumption simplifies our problem and nevertheless allows us to capture the dual effect of fatigue in our model. More precisely, fatigue affects the problem both through search costs, i.e., through the term αt, and through its effect on the continuation value $W_j(\bar{S}, t+1, y)$, captured in a reduced-form way in Assumption 1 through the term δt.[28]

Given Assumption 1, the problem in Equation (6) can be rewritten as

$$V(j^*, t, y) = \max_{stop,\, now,\, later} \left\{ y,\; -c_{j^*0} - (\alpha - \delta)\, t + W_{j^*}(y),\; \beta \left[-c_{j^*0} + W_{j^*}(y) \right] \right\}. \tag{7}$$

The problem stated in Equation (7) is straightforward to solve. In particular, for j^*, the product with the largest reservation utility at (t, y), the search rules are given by

- search j^* now if $c_{j^*0} \;\leq\; W_{j^*}(y) - y - (\alpha - \delta)\, t$ and $c_{j^*0} \;\leq\; W_{j^*}(y) - \frac{(\alpha-\delta)t}{1-\beta}$;
- search j^* later if $c_{j^*0} \;\leq\; W_{j^*}(y) - \frac{y}{\beta}$ and $c_{j^*0} \;>\; W_{j^*}(y) - \frac{(\alpha-\delta)t}{1-\beta}$;
- stop searching if $c_{j^*0} \;>\; W_{j^*}(y) - y - (\alpha - \delta)\, t$ and $c_{j^*0} \;>\; W_{j^*}(y) - \frac{y}{\beta}$.

Recall that $W_j(y) = yF_j(y) + \int_y^\infty u dF_j(u)$. As shown in Weitzman (1979), the continuation value $W_j(y)$ is a positive, continuous, and monotonically increasing function of y with $\partial W_j(\cdot)/\partial y = F_j(y) \geq 0$. Also, $W_j(y) - y$ is a positive, continuous, and monotonically decreasing function of y. Therefore, there exists a unique solution z_j to the equation

$$c_{j0} = W_j\left(z_j\right) - z_j, \tag{8}$$

which represents the reservation utility of an option (Weitzman (1979)). Similarly, there exist unique solutions to each of the inequalities describing the consumer search rules above. More precisely, because search costs $c_{j0} > 0$ and the value of $(\alpha - \delta)\, t$ are constant in y, while $W_j(y) - y$ is positive, continuous, and monotonically decreasing in y, there exists a unique solution $z_j^1(t)$ to the equation

$$c_{j0} = W_j\left(z_j^1(t)\right) - z_j^1(t) - (\alpha - \delta)\, t. \tag{9}$$

Also, whenever $\beta \neq 0$, there exists a unique solution z_j^2 to the equation

$$c_{j0} = W_j\left(z_j^2\right) - \frac{z_j^2}{\beta} \tag{10}$$

[28] Without data that would allow us to separately identify the effect of fatigue through search costs or the continuation value, we cannot test the specific functional form on the term δt, which leads us to assume linearity. More details can be found in Section 6.3.

since $W_j(y) - y$ is continuous and monotonically decreasing in y. Finally, whenever $\beta < 1$, there exists a unique solution $z_j^3(t)$ to the equation

$$c_{j0} = W_j\left(z_j^3(t)\right) - \frac{(\alpha - \delta)t}{1-\beta}. \tag{11}$$

Based on these results, **our proposed solution** is given by:

If $t > 0$:

1. **Selection rule:** order alternatives in decreasing order of reservation utilities z_j (as defined by Weitzman 1979).

2. **Search rules:** if j is the alternative with the maximum reservation utility among alternatives not yet searched $\bar{S}$, then given (t, y)

   - search j *now* if $\max\left\{z_j^1(t), z_j^2\right\} \geq y$ and $z_j^3(t) < y$;
   - search j *later* if $\max\left\{z_j^1(t), z_j^2\right\} \geq y$ and $z_j^3(t) \geq y$;
   - *stop* searching if $z_j^1(t) < y$ and $z_j^2 < y$.

3. **Choice rule:** upon stopping, purchase the option with the largest realized value among those searched, $y = \max_{j \in S} u_j$, or choose the outside option of not purchasing.

When $t = 0$, our model reduces to the one in Weitzman (1979), in which the consumer searches an option j now if $z_j \geq y$ and stops otherwise. Note that, if $t = 0$, i.e., fatigue does not affect search decisions, the consumer searches more options than in the case of $t > 0$. To see this, note that $z_j^1(t) < z_j$, $\forall t > 0$ and $z_j^2 < z_j$ since $\beta < 1$. Therefore, the condition for search with fatigue (requiring at least that $\max\left\{z_j^1(t), z_j^2\right\} \geq y$) is less likely to hold than the search condition ($z_j \geq y$) in the case with no fatigue, leading to fewer searches. Similarly, larger fatigue levels affect the number of searched options.

6 Empirical Application

6.1 Empirical Model

We take the theoretical model presented in the previous section to data using the following empirical specification: we model consumers as searching across websites (e.g., zalando.nl or nike.com). Specifically, consumer $i = 1, \ldots, N$ seeks to purchase from website $j = 1, \ldots, J$ or to choose the outside option of not purchasing (denoted by $j = 0$). Consumer i's utility for website j is given by

$$\begin{aligned} u_{ij} &= v_{ij} + \epsilon_{ij} \\ &= w_j + \gamma X_{ij} + \eta_{ij} + \epsilon_{ij} \end{aligned} \tag{12}$$

where v_{ij} denotes the information the consumer has about a website before searching it, while ϵ_{ij} denotes the information she searches for. Before searching, the consumer knows individual websites' values which are denoted by website intercepts w_j. In each subcategory, we estimate separate website intercepts for the 10 most searched websites (accounting for approximately 65% of clicks in each subcategory) and group all other websites into a composite reference website called "Other." Although we scraped prices and other product features from the URLs provided in our data, these features do not generally vary over time or across consumers in the fashion industry (see also Section 4.3). Therefore, after controlling for website intercepts, the effects of such features cannot be separately estimated. Nevertheless, we include two additional controls X_{ij} in the utility function: (i) the number of times the consumer has previously searched a given website (across all product subcategories) to measure her loyalty for and knowledge of a website, and (ii) an indicator for whether the consumer visited a price discount page to partially capture her price sensitivity.

Next, η_{ij} is the part of the utility that is observed by the consumer (prior and post search), but not the researcher (neither prior nor post search). It captures deviations from website features that the consumer may be aware of before starting her search. Consumers search sequentially to resolve uncertainty about their match values ϵ_{ij}, i.e., consumers do not observe ϵ_{ij} prior to search but they do post search. The researchers neither observes ϵ_{ij} prior nor post search. ϵ_{ij} captures everything the consumer learns by visiting the website, e.g., available (actual) product styles, sizes, colors, customer reviews, photos, etc. η_{ij} and ϵ_{ij} are both standard normally distributed. The outside option does not require searching and is modeled as $u_{i0} = q_0 + \eta_{ij}$, where q_0 is an intercept denoting the value of not purchasing.

Searching to resolve uncertainty about ϵ_{ij} is costly to consumers. Search costs (per search) are given by

$$\begin{aligned} c_i &= c_0 + \alpha_i t \\ &= \exp(\kappa_0) + \exp(\lambda_0 + \lambda_1 Age_i)\, t \end{aligned} \tag{13}$$

where c_0 are baseline search costs and $t > 0$ captures the number of searches performed without a break. To ensure that search costs are positive, we operationalize search costs as exponential functions consistent with prior work (e.g., Honka 2014; Chen and Yao 2017; Ursu 2018). Lastly, consistent with our results from Table 5, we use age as a shifter of fatigue.

6.2 Estimation

We use the search rules from Section 5.4 to construct the likelihood of consumers' search and purchase decisions. These rules translate into the following restrictions on preferences, search costs, and fatigue parameters. Suppose a consumer i searched a number of options s of the total J websites available and she chose j after stopping her search (including the outside option). With a slight abuse of notation, order websites by their reservation utilities and let n denote the website with the nth largest reservation utility. Also, let t_n denote the number of websites the consumer searched since the previous gap and before searching n.

Since consumers search websites in a decreasing order of their reservation utilities, according to the *selection rule*, it must be that

$$z_{in} \geq \max_{k=n+1}^{J} z_{ik} \qquad \forall n \in \{1, \ldots, J-1\}. \tag{14}$$

For the first searched website (for which $t_1 = 0$), the *selection rule* additionally requires that its reservation utility exceeds the utility of the outside option, i.e., $z_{i1} \geq u_{i0}$. Since all consumers in our data search at least once, consistent with prior work (e.g., Honka 2014; Honka and Chintagunta 2017), we assume that the first search is free (i.e., we do not impose this constraint in empirical application, but allow for the possibility of no search in our Monte Carlo simulation in Section 6.4).

After searching the first website, $t_n > 0 \;\; \forall n > 1$ and the *search rules* describe consumer behavior. For searched website n we know that

$$\max\left\{z_{in}^1(t_n), z_{in}^2\right\} \geq \max_{k=0}^{n-1} \; u_{ik} \qquad \forall n \in \{2, \ldots, s\}. \tag{15}$$

What separates our model from previous work is that we additionally capture a consumer's decision of when to search an option (with or without a break). In particular, all websites, except the first searched one, may be searched with or without a gap. Thus, if n was searched without a break, i.e., $t_n = t_{n+1} - 1$, we know that

$$z_{in}^3(t_n) < \max_{k=0}^{n-1} \; u_{ik} \qquad \forall n \in \{2, \ldots, s\}, \tag{16}$$

while, if n was searched after a break, i.e., if $t_n \neq t_{n+1} - 1$, it must be that

$$z_{in}^3(t_n) \geq \max_{k=0}^{n-1} \; u_{ik} \qquad \forall n \in \{2, \ldots, s\}. \tag{17}$$

For all options m that were not searched, it must be that

$$z_{im}^1(t_m) \;<\; \max_{k=0}^{s} \; u_{ik} \qquad \forall m \in \{s+1, \ldots, J\} \tag{18}$$

$$z_{im}^2 \;<\; \max_{k=0}^{s} \; u_{ik} \qquad \forall m \in \{s+1, \ldots, J\}. \tag{19}$$

with $t_m = t_s + 1$.

Finally, consistent with the *choice rule*, if the consumer chooses j (including the outside option), her utility from this choice exceeds that of all searched websites, i.e.,

$$u_{ij} = \max_{k=0}^{s} \; u_{ik} \qquad \forall j \in \{0, 1, \ldots, s\}. \tag{20}$$

If consumers search using the rules described above, then they make search, search gap, and purchase decisions jointly. Thus, the probability of observing a certain outcome in the data for consumer i is characterized by the joint probability of Equations (14)-(20) holding. This probability is given by

$$L_i = Pr(\text{Selection rule}_i, \text{ Search rule}_i, \text{ Choice rule}_i). \tag{21}$$

Because consumers make these decisions jointly, the likelihood function does not have a closed-form solution. We use a simulated maximum likelihood (SMLE) approach to estimate the parameters of the model. In choosing the simulation method, we follow McFadden (1989), Honka (2014), Honka and Chintagunta (2017), Ursu (2018), and Ursu, Wang, and Chintagunta (2020) and use the logit-smoothed AR simulator. The implementation details are discussed in Web Appendix B.

An advantage of our proposed method lies in its ease of estimation due to its similarity to the Weitzman (1979) model: consumers search in decreasing order of reservation utilities z_j and also make search and purchase decisions based on threshold values of the best observed alternative so far.

The main difference consists of computing the values of $\left[z_j^1(t), z_j^2, z_j^3(t)\right]$ in addition to that of z_j. We describe how to compute $\left[z_j^1(\cdot), z_j^2, z_j^3(\cdot)\right]$ in Web Appendix C.

We estimate our model using data from the two most commonly purchased subcategories in our data: "shirts, tops, & blouses" and "shoes" to demonstrate that our results are not limited to a specific subcategory. Details on the construction of the estimation samples are provided in Web Appendix A.

6.3 Identification

The set of parameters to be estimated is composed of the utility parameters w_j and γ, search cost c_0 (parameterized by κ_0), search fatigue α (parameterized by λ_0 and λ_1), continuation fatigue δ, and the discount factor β. As is well-known, the discount factor is not typically identified in dynamic discrete choice models without further restrictions (Rust 1994; Magnac and Thesmar 2002). This is the case in our model as well. Thus, consistent with the previously mentioned paper, we fix the discount factor to 0.95.[29] Further, with data on searches, search gaps, and purchase decisions, we can only recover the difference between search fatigue α and continuation fatigue δ. We therefore set $\delta = 0$ in the estimation and emphasize that the estimated search fatigue parameter is an underestimate of the true value of search fatigue. Given these considerations, the set of parameters to be estimated is $\theta = (w_j, \gamma, \kappa_0, \lambda_0, \lambda_1)$.

As standard in consumer search models, utility parameters are identified from search and purchase frequencies observed in the data. For example, websites that are searched and purchased more frequently will have a larger estimated value. Also, variation in the frequencies with which consumers have previously visited websites and whether they visit price discount pages identify γ.

Similarly, as in prior work, search costs do not affect purchase decisions and are identified from the number of websites that consumers search. More precisely, the search rules impose an upper and a lower bound on search cost c_0 that must have made it optimal for the consumer to perform a certain number of searches. These search rules, however, only recover a range of search costs. The level of search costs is pinned down by the functional form and the distribution of the utility function that dictate the reservation utility expressions in Equations (C1), (C2), and (C4).

By observing search gaps, i.e., consumers' decisions of *when* to search each website, we can additionally identify consumer fatigue levels. In other words, conditional on an observed number of searches, search gaps identify the fatigue parameter λ_0. Finally, deviations in search gaps across

[29] We estimate our model with the discount factor set to 0.90 as a robustness check. The results are displayed in Table D-2 in Web Appendix D and show fatigue and search cost parameters that are slightly larger than those in our main specification, but overall robust.

consumers attributable to age identify λ_1.

6.4 Monte Carlo Simulation

To show that our estimation procedure can recover the model parameters, we perform the following Monte Carlo simulation exercise. We generate a data set of 5,000 consumers making choices among five options (one outside option and four websites). Consumers value each website differently. Search costs have two components: a baseline level of search costs and fatigue. The true values of the utility and both search cost parameters are similar to those from a preliminary estimation of our model.

For estimation, we follow the steps described in Section 6.2 and use 200 draws from the distribution of the utility error terms (both η_{ij} and ϵ_{ij}) for each consumer-website combination. We simulate 50 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 7. In column (i), we present the true parameters; in column (ii), we show the mean of the estimated parameters across the 50 simulations and the standard deviation of the mean across these simulations. Our proposed estimation procedure recovers the model parameters well. In addition, in column (iii), we also report results from the Weitzman model that ignores search gaps. We obtain these results by estimating the model on the same data, but assuming no gaps occurred in the data.[30] We find that the Weitzman model performs well in recovering the true utility parameters, but overestimates baseline search costs. We discuss the estimation bias observed here after presenting our estimation results in the next section.[31]

==========

Insert Table 7 about here

==========

7 Results

7.1 Model Estimates

We show the estimation results for the "shirts, tops, & blouses" and "shoes" subcategories in Table 8. For each subcategory, the first two columns display results from our model that accounts for search

[30] For details on how the Weitzman model was estimated, see Web Appendix B.

[31] The difference in log-likelihood values reported comes from the difference in the likelihood function and the number of parameters estimated. As we explain in Section 7.3 below, log-likelihood values cannot be used to compare model fit.

gaps, while the third column reports results from the Weitzman model that ignores search gaps.[32] The results indicate that consumers derive positive utility from the outside option, consistent with the empirical observation that most consumers do not make a purchase. Next, we find that Zalando, the largest online retailer in the Netherlands, is among the most preferred websites in both subcategories, together with C&A in the "shirts, tops, & blouses" subcategory and Schurrman Shoenen in the "shoes" subcategory. As expected, previous visits to a website increase a consumer's utility for that website. Although we are not able to account for the effect of prices, we show consumers' price sensitivity by reporting that consumers who visited the price discount page (e.g., the "sale" or "clearance" page) of a website derive higher utility.

==========

Insert Table 8 about here

==========

The baseline search cost estimate is relatively small compared to the fatigue estimate. To put it differently, fatigue has a large effect on consumer search decisions, equivalent to increasing baseline search costs at least tenfold with every searched website.[33] Also, the larger number of search gaps in the "shoes" subcategory reveals a larger fatigue level (as a proportion of baseline search costs) for consumers in that subcategory (the ratio of fatigue to baseline search costs is 14 in the "shirts, tops, & blouses" subcategory and it is 18 in the "shoes" subcategory). Finally, consistent with our reduced-form results in Table 5), older consumers have even larger fatigue costs.

7.2 Comparison with Weitzman (1979)

By comparing our results to those from the Weitzman model that ignores search gaps, we provide insights into the estimation bias that arises when search gaps are ignored.[34] In particular, both our estimation and our simulation results show that the utility parameter estimates are similar in our model and the Weitzman model, but that baseline search costs are overestimated in the latter one.[35] The intuition behind this difference is as follows: recall that the Weitzman model ignores search gaps and assumes that search costs are independent of the number of previously searched options. Thus,

[32] We also estimated a version of our model in which we set $\alpha = 0$ and assume no gaps occurred in the data. The results are very similar to those from the Weitzman model and available from the authors upon request.

[33] Calculation follows after dividing the fatigue constant by the baseline search cost parameter, e.g., in the first column, $\exp(-2.4436)/\exp(-5.0740) \approx 14$.

[34] We use the term "bias" to denote the difference in estimates based on our results.

[35] Note that, if we had estimated the Weitzman model at the session level rather than at the spell level, similar to some of the literature that does not observe search gaps (e.g., Ursu 2018), then our results would have shown an even larger difference.

when estimated on the same data set as a model in which fatigue affects search costs, the Weitzman model rationalizes the same number of searched products by inflating baseline search costs.

Although the differences in intercept estimates for a particular website in our and the Weitzman model are small, theoretically the intercepts for a particular website in both models are different. Whether this difference in intercepts is positive or negative depends on whether a particular website is predominantly searched before or after search gaps. In our empirical application, the Weitzman model provides larger estimates of website intercepts than our model. This finding is consistent with the observation that many websites are searched after a break in our data.

Our model makes two changes to the Weitzman framework: (i) allows for search gaps; and (ii) models the effect of fatigue on search costs. To better isolate the effect of each change on parameter estimates, we also estimate a (variation of the) Weitzman model with increasing search costs (due to fatigue) but without search gaps, i.e., we only make one change to the Weitzman framework. To the best of our knowledge, such a variation of the Weitzman model has not been studied by previous literature. As in our problem, there is no known optimal search rule. However, in Web Appendix D, we describe how the solution we developed for our model can be used to derive an optimal search rule for the Weitzman model with increasing search costs.

The estimation results are displayed in Table D-1 in Web Appendix D. Not accounting for search gaps leads to an overestimation of the baseline search cost, although the bias is smaller when at least fatigue is taken into account. In addition, a Weitzman model with increasing search cost underestimates the importance of fatigue compared to a model that accounts for search gaps. The intuition for this results is as follows: the number of previously searched options when breaks are allowed is equal or smaller than the number of previously searched options when breaks are not allowed. This makes it seem like fatigue plays a smaller role in the consumer's decision to search when breaks are not allowed, since the consumer chooses to continue searching despite the large fatigue level, resulting in a smaller fatigue estimate. In contrast, accounting for the fact that fatigue caused the consumer to take a break from searching, rather than not only increased search costs, reveals the larger importance of fatigue.

7.3 Model Fit

Because of the difference in the likelihood functions and in the number of parameters, we cannot rely on the log-likelihood reported in Table 8 to compare our model to the Weitzman model. Instead, to

understand which model better captures consumer behavior, we calculate the root mean squared error (RMSE) for the three decisions consumers make in our model (search gaps, searches, and purchases). For these calculations, we use the first set of estimation results reported in Table 8 for each subcategory to obtain the RMSE for our model and the estimation results for Weitzman model. The RMSE results are displayed in Table 9.[36]

==========

Insert Table 9 about here

==========

Our model recovers the number of search gaps well, while the Weitzman model (by construction) cannot predict any such decisions. When comparing the search shares of each website, our model more accurately predicts which websites consumers search more frequently. Finally, although the utility estimates in the two models are very similar, the Weitzman model better predicts market shares. In part, this difference also arises from the fact that our model explains three rather than only two decisions that consumers make. In a similar vein, a simple choice model that focuses on explaining only one decision would better predict market shares than the Weitzman model that describes two decisions (e.g., in our data, a conditional logit model explaining a consumer's purchase decision across websites results in market share RMSE of 12 and 9 in the two subcategories, respectively).

8 Counterfactuals

In this section, we present results from two counterfactuals. Our analysis should be interpreted as capturing short-run effects of changes in consumer primitives.

8.1 Effects of Reducing Fatigue

Managers have long sought ways to reduce consumers' burden from shopping and the fatigue resulting from it. Reasons contributing to consumer fatigue include, e.g., access to a plethora of information sources (expert reports, social media, customer reviews and ratings, opinion blogs, etc.), the choice among an ever growing number of products with different attributes, and the pressure to choose

[36] The RMSE for the Weitzman model with increasing search costs are as follows: for' 'shirts, tops, & blouses," 58 and 44 for search and market shares, respectively. For "shoes," 71 and 52 for search and market shares, respectively. RMSE for search gaps is the same as for the Weitzman model.

the best deal when receiving numerous promotions.[37] In our setting, we showed in Section 4 that a website's loading speed and readability also contribute to consumer fatigue. Thus improvements in website design to increase readability or an increase in a website's loading speed represent two potential avenues to reduce consumer fatigue.

Here, we quantify the value of reducing fatigue and investigate which brands benefit relatively more from a reduction in fatigue than others. To accomplish these goals, for each of the two fashion subcategories, we employ our model and coefficient estimates to simulate consumer decisions regarding searches, search gaps, and purchases in three scenarios (holding everything else constant): (i) when fatigue is reduced by 50%; (ii) when fatigue is reduced by 90%; and (iii) when baseline search costs are reduced by 50%. Consumers' simulated decisions in these scenarios are then compared to the current setting in which no such change occurs. To integrate over the distribution of unobserved utility shocks in the model, we repeat the simulation 50 times and report the mean results.

We present our findings in Table 10. The effects are similar in both fashion subcategories although they are larger in the "shoes" subcategory for which we found higher fatigue levels. In particular, we find that decreasing fatigue by 50% increases the number of searched websites by $1-4\%$, increases transactions by $0.5-1.2\%$, and lowers search gaps by at least 11%. Interestingly, this effect is larger than that of decreasing baseline search costs by 50%, a policy that most prior work focused on (e.g., Seiler 2013, Honka 2014, Moraga-González, Sándor, and Wildenbeest 2018, Yavorsky, Honka, and Chen 2020). Also, in contrast to the effect of reducing fatigue, decreasing baseline search costs increases search gaps: the more options the consumer searches, the higher the chances of her taking a break from searching, and thus the more gaps. When fatigue is decreased by a factor of 10, these results are magnified.

==========

Insert Table 10 about here

==========

We also investigate which websites benefit the most from a reduction in fatigue. Our results for both subcategories are displayed in Figure 2. For this analysis, we utilize our results from the case in which fatigue is decreased by 90% to better highlight the magnitude of the effects. To present our results, in each figure, we order the top 10 websites by their size, i.e., by their search or market shares,

[37] For more details, see http://business.time.com/consumer-fatigue-shopping-has-never-been-easier-or-as-mentally-exhausting/ or https://hbr.org/death-by-information-overload.

respectively, as per Table 3. Not surprisingly, all websites benefit from a reduction in fatigue, but smaller and less popular websites benefit relatively more than larger and more popular websites. The reason for this finding is that a reduction in fatigue leads to additional searches and purchases that would not have occurred for smaller websites if fatigue levels had been higher.

==========

Insert Figure 2 about here

==========

8.2 When Breaks Do Not Decrease Fatigue

Although search costs due to fatigue reset after a break in our model, it is not a forgone conclusion that consumers can always lower their fatigue levels during a break. On the contrary, several recent articles talk about the consequences of consumers being constantly stimulated by marketing activities[38] and of the Covid-19 pandemic keeping consumer fatigue levels high.[39] In such cases, a managerially relevant question to ask is how an inability to decrease fatigue levels affects consumer decisions.

We answer this question in our second counterfactual. Our simulation procedure is similar to the one described for the first counterfactual; however, the analyzed scenario differs. More precisely, we simulate consumer search and purchase decisions for the case in which the fatigue level α does not reset to zero during a break. In this case, consumers always prefer to search now rather than to delay their search, since delaying leads to a discounted value of the same expected utility as that of searching now. Therefore, the problem consumers solve in this case reduces to that of deciding whether to search now or whether to stop, and the optimal decision rule follows after inspecting Equations 2 and 3 without the option to search later. In particular, the consumer continues to search website j if $z_j^1(t) \geq y$ and stops otherwise.

The results are displayed in Table 11. Not being able to reset fatigue during a break leads to a significant reduction in the number of searched and purchased products: searches decrease by approximately 20% and purchases by more than 6%. Both of these effects are large, e.g., in comparison to those obtained when reducing fatigue.

[38] Information is available in the following articles: https://hbr.org/to-keep-your-customers-keep-it-simple, https://www.nytimes.com/do-you-suffer-from-decision-fatigue.

[39] See for example https://www.delish.com/grocery-shopping-brand-loyalty-reason/, https://hbr.org/how-to-combat-zoom-fatigue, https://hbr.org/coping-with-fatigue-fear-and-panic-during-a-crisis.

==========

Insert Table 11 about here

==========

Although all websites suffer in this setting, different websites are affected differently. We display the results for both subcategories in Figure 3. Here, we again order websites by their their search or market shares, respectively, as per Table 3. We find that larger and more popular websites are hurt less when consumers cannot effectively take a break from searching to reduce their fatigue levels. In other words, website loyalty becomes stronger for the most popular websites at the expense of smaller ones. Consumers prefer to search and choose more often websites they are familiar with.

==========

Insert Figure 3 about here

==========

9 Conclusion

In this paper, we document that consumers frequently decide to take breaks during their search process and provide model-free evidence that such breaks are related to fatigue. To rationalize such behavior, we develop a model of sequential search that extends the Weitzman (1979) framework and allows for search gaps due to fatigue. We quantify the effect of fatigue on consumer search and purchase decisions and show the possible estimation bias in search costs when search gaps are ignored. Finally, we quantify the managerial value of reducing consumer fatigue through a series of counterfactuals.

There are several potentially useful extensions to our approach. First, it would be interesting to explore other drivers of search gaps across a variety of product categories. Our model can provide a starting point for formalizing the mechanism behind search gaps in such settings. Second, future work could consider the important aspects of consumer learning and forgetting and their role in affecting search gaps. Another possible extension would look at what explains the length of search gaps, in addition to their occurrence. Also, extending our model to account for brand specific fatigue could provide another interesting avenue for researchers to explore. We leave these and other related topics to future research.

References

Abaluck, J. and J. Gruber (2013), "Choice Inconsistencies among the Elderly: Evidence from Plan Choice in the Medicare Part D Program," *American Economic Review*, 101 (4), 1180–1210.

Anderson, S. and R. Renault (2018), "Firm Pricing with Consumer Search," in L. Corchón and M. Marini (Editors), "Handbook of Game Theory and Industrial Organization," Edward Elgar, volume 2, 177–224.

Arbatskaya, M. (2007), "Ordered Search," *RAND Journal of Economics*, 38 (1), 119–126.

Armstrong, M., J. Vickers, and J. Zhou (2009), "Prominence and Consumer Search," *RAND Journal of Economics*, 40 (2), 209–233.

Baye, M., J. Morgan, and P. Scholten (2006), "Information, Search, and Price Dispersion," *Handbook on Economics and Information Systems*, 1, 323–377.

Bednar, S. and K. Rouse (2019), "Does Taking a Break Make Students (Un)Productive?" Working Paper.

Brown, D. and J. Smith (2013), "Optimal Sequential Exploration: Bandits, Clairvoyants, and Wildcats," *Operations Research*, 61 (3), 644–665.

Carlin, B. and F. Ederer (2019), "Search Fatigue," *Review of Industrial Organization*, 54 (3), 485–508.

Carpenter, S. and C. Yoon (2011), "Aging and Consumer Decision Making," *Annals of the New York Academy of Sciences*, 1235 (1), E1–E12.

Cavallo, A. (2017), "Are Online and Offline Prices Similar? Evidence from Large Multi-Channel Retailers," *American Economic Review*, 107 (1), 283–303.

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

Dang, C., R. Ursu, and P. Chintagunta (2020), "Search Revisits," Working Paper.

De los Santos, B. and S. Koulayev (2017), "Optimizing Click-Through in Online Rankings with Endogenous Search Refinement," *Marketing Science*, 36 (4), 542–564.

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.

Dhar, R. (1997), "Context and Task Effects on Choice Deferral," *Marketing Letters*, 8 (1), 119–130.

Ellison, G. and S. Ellison (2009), "Search, Obfuscation, and Price Elasticities on the Internet," *Econometrica*, 77 (2), 427–452.

Ellison, G. and A. Wolitzky (2012), "A Search Cost Model of Obfuscation," *RAND Journal of Economics*, 43 (3), 417–441.

Gittins, J., K. Glazebrook, and R. Weber (2011), *Multi-Armed Bandit Allocation Indices*, John Wiley and Sons Inc., 2nd edition.

Greenleaf, E. and D. Lehmann (1995), "Reasons for Substantial Delay in Consumer Decision Making," *Journal of Consumer Research*, 22 (2), 186–199.

Harada, C., M. Natelson Love, and K. Triebel (2013), "Normal Cognitive Aging," *Clinics in Geriatric Medicine*, 29 (4), 737–752.

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

Honka, E. and P. Chintagunta (2017), "Simultaneous or Sequential? Search Strategies in the U.S. Auto Insurance Industry," *Marketing Science*, 36 (1), 21–40.

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

Honka, E., A. Hortascu, and M. Wildenbeest (2019), "Empirical Search and Consideration Sets," in JP Dube and P Rossi (Editors), "Handbook on the Economics of Marketing," North Holland.

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

Kim, S., D. Goldstein, L. Hasher, and R. Zacks (2005), "Framing Effects in Younger and Older Adults," *The Journals of Gerontology. Series B, Psychological Sciences and Social Sciences*, 60 (4), 215–218.

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

Lambert-Pandraud, R., G. Laurent, and E. Lapersonne (2005), "Repeat Purchasing of New Automobiles by Older Consumers: Empirical Evidence and Interpretations," *Journal of Marketing*, 69 (2), 97–103.

Levav, J., M. Heitmann, A. Herrmann, and S. Iyengar (2010), "Order in Product Customization Decisions: Evidence from Field Experiments," *Journal of Political Economy*, 118 (2), 274–299.

Magnac, T. and D. Thesmar (2002), "Identifying Dynamic Discrete Choice Processes," *Econometrica*, 70 (2), 801–816.

McFadden, D. (1989), "A Method of Simulated Moments for Estimation of Discrete Response Models without Numerical Integration," *Econometrica*, 57 (5), 995–1026.

Moraga-González, J., Z. Sándor, and M. Wildenbeest (2018), "Consumer Search and Prices in the Automobile Market," Working Paper.

Morgan, P. and R. Manning (1985), "Optimal Search," *Econometrica*, 53 (4), 923–944.

Novemsky, N., R. Dhar, N. Schwarz, and I. Simonson (2007), "Preference Fluency in Choice," *Journal of Marketing Research*, 44 (3), 347–356.

Ozdenoren, E., S. Salant, and D. Silverman (2012), "Willpower and the Optimal Control of Visceral Urges," *Journal of the European Economic Association*, 10 (2), 342–368.

Peters, E. (2010), "Age Related Changes in Decision Making," in A. Drolet, N. Schwarz, and C Yoon (Editors), "The Aging Consumer: Perspectives from Psychology and Economics," Taylor and Francis, 75–101.

Petrikaite, V. (2018), "Consumer Obfuscation by a Multiproduct Firm," *RAND Journal of Economics*, 49 (1), 206–223.

Rust, J. (1994), "Structural Estimation of Markov Decision Processes," in R. Engle and D. McFadden (Editors), "Handbook of Econometrics," Amsterdam: North-Holland, volume 4, 3081–3143.

Schmittlein, D., D. Morrison, and R. Colombo (1987), "Counting Your Customers: Who Are They and What Will They Do Next?" *Management Science*, 33 (1), 1–24.

Seiler, S. (2013), "The Impact of Search Costs on Consumer Behavior: A Dynamic Approach," *Quantitative Marketing and Economics*, 11 (2), 155–203.

Stigler, G. (1961), "The Economics of Information," *Journal of Political Economy*, 69 (3), 213–225.

Stiglitz, J. (1987), "Competition and the Number of Firms in a Market: Are Duopolies More Competitive than Atomistic Markets?" *Journal of Political Economy*, 95 (5), 1041–1061.

Tanius, B., S. Wood, Y. Hanoch, and T. Rice (2009), "Aging and Choice: Applications to Medicare Part D," *Judgement and Decision Making*, 4 (1), 92–101.

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

Ursu, R. and D. Dzyabura (2020), "Retailers' Product Location Problem with Consumer Search," *Quantitative Marketing and Economics*, 18 (3), 125–154.

Ursu, R., Q. Wang, and P. Chintagunta (2020), "Search Duration," *Marketing Science*, 39 (5), 849–871.

Vohs, K., R. Baumeister, B. Schmeichel, J. Twenge, N. Nelson, and D. Tice (2008), "Making Choices Impairs Subsequent Self-Control: A Limited-Resource Account of Decision Making, Self- Regulation, and Active Initiative," *Journal of Personality and Social Psychology*, 94 (5), 883–898.

Weitzman, M. (1979), "Optimal Search for the Best Alternative," *Econometrica*, 47 (3), 641–654.

Whittle, P. (1980), "Multi-Armed Bandits and the Gittins Index," *Journal of the Royal Statistical Society Series B (Methodological)*, 42 (2), 143–149.

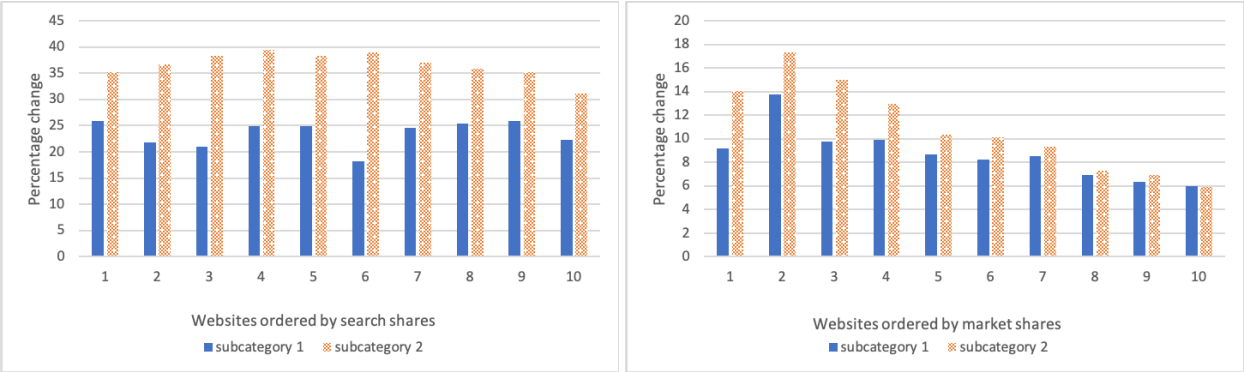
Wu, C., H. Che, T. Chan, and X. Lu (2015), "The Economic Value of Online Reviews," *Marketing Science*, 34 (5), 739–754.

Yavorsky, D., E. Honka, and K. Chen (2020), "Consumer Search in the U.S. Auto Industry: The Role of Dealership Visits," *Quantitative Marketing and Economics*, forthcoming.

Figures and Tables

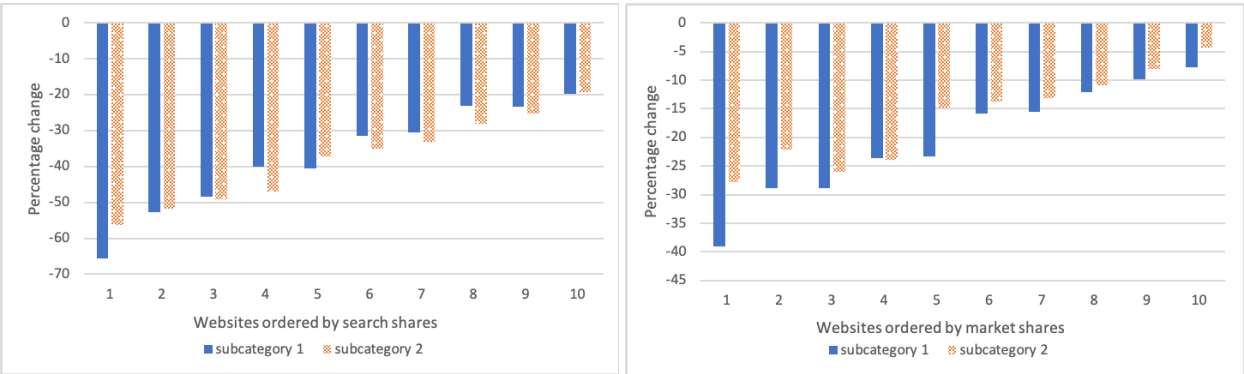
Figure 1: Example of a Search Process with Definitions

Figure 2: Effects of Lower Fatigue Levels by Brand



(a) Search shares **(b) Market shares**

Figure 3: Effects of No Breaks by Brand



(a) Search shares **(b) Market shares**

Table 1: Session and Spell Characteristics

	Mean	Median	St. Dev.	Min	Max
Session: All Clicks					
Number of Clicks	189.07	116.00	250.85	1	9,108
Number of Visited Websites	29.38	20.00	28.44	1	309
Duration (in minutes)	95.58	64.25	106.11	0	2,043
Session: Fashion Clicks					
Number of Clicks	10.52	3.00	26.34	1	2,407
Number of Visited Websites	1.73	1.00	1.44	1	29
Duration (in minutes)	5.59	1.03	12.21	0	508
Number of Subcategories	1.14	1.00	1.29	0	9

Table 2: Most Popular Fashion Websites and Product Subcategories

	Popular Websites		Ranked Subcategories
Ordered by	*Search share*	*Market share*	*Market share*
	zalando.nl	zalando.nl	Shirts, tops, & blouses
	hm.com	hm.com	Shoes
	c-and-a.com	c-and-a.com	Pants & jeans
	debijenkorf.nl	your-look-for-less.nl	Underwear
	missetam.nl	esprit.nl	Dresses & skirts
	your-look-for-less.nl	debijenkorf.nl	Children's clothes
	vente-exclusive.com	missetam.nl	Jackets & vests
	esprit.nl	vente-exclusive.com	Accessories
	vanharen.nl	hunkemoller.nl	
	schuurman-shoenen.nl	vanharen.nl	

Table 3: Top 10 Websites for the "Shirts, Tops, &Blouses" and "Shoes" Subcategories

	"Shirts, tops, & blouses"		"Shoes"	
Ordered by	*Search share*	*Market share*	*Search share*	*Market share*
	c-and-a.com	zalando.nl	zalando.nl	zalando.nl
	debijenkorf.nl	hm.com	schuurman-shoenen.nl	vanharen.nl
	zalando.nl	your-look-for-less.nl	vanharen.nl	adidas.com
	hm.com	c-and-a.com	adidas.com	debijenkorf.nl
	aboutyou.com	esprit.nl	spartoo.nl	nelson.nl
	esprit.nl	peterhahn.nl	nike.com	nike.com
	your-look-for-less.nl	aboutyou.com	omoda.nl	omoda.nl
	msmode.nl	debijenkorf.nl	nelson.nl	schuurman-shoenen.nl
	peterhahn.nl	jbfo.nl	debijenkorf.nl	spartoo.nl
	jbfo.nl	msmode.nl	ziengs.nl	ziengs.nl

Table 4: Search Gaps within Fashion Subcategories

	Shirts, Tops & Blouses	Shoes	Pants & Jeans	Under-wear	Sweaters	Dresses & Skirts	Children's Clothes	Jackets & Vests	Acces-sories
Proportion of Spells with ≥ 1 Search Gap	0.50	0.56	0.44	0.36	0.37	0.39	0.45	0.38	0.41
Av. Number of Search Gaps if ≥ 1 Search Gap	2.87	5.14	2.65	2.21	2.43	2.90	2.77	2.54	2.81
Av. Length of Search Gaps (in Days)	7.39	5.99	7.92	8.89	8.24	7.32	8.28	7.60	7.72
Av. Length of Spell (in Days)	10.58	17.30	9.21	7.03	7.42	8.22	10.45	7.29	8.86
Av. Time between Spells (in Days)	11.73	14.23	14.12	19.70	12.27	15.86	16.29	19.64	19.99

Table 5: Effects of Fatigue Proxies on Number of Search Gaps

	Dependent variable:			
	Number of search gaps in a spell[a]			
	(i)	(ii)	(iii)	(iv)
Subcategory	"Shirts, top and blouses"		"Shoes"	
Consumer characteristics				
Age	0.0037***	0.0037***	0.0095***	0.0103***
	(0.0008)	(0.0009)	(0.0009)	(0.0011)
Female indicator	0.1297***	0.1444***	-0.0188	-0.0358
	(0.0270)	(0.0316)	(0.0301)	(0.0363)
Time variables				
Evening dummy	-0.0691*	-0.0722*	-0.0753*	-0.0959*
	(0.0285)	(0.0344)	(0.0345)	(0.0430)
Weekend dummy	0.0419	0.0643	0.0308	0.0979*
	(0.0312)	(0.0378)	(0.0362)	(0.0457)
Website characteristics				
Speed score		-0.0020*		-0.0004
		(0.0010)		(0.0009)
Readability (SMOG)		-0.0071		0.0405***
		(0.0108)		(0.0047)
Number of images		-0.0002		0.0001
		(0.0002)		(0.0002)
Number of words		-0.0000		-0.0000**
		(0.0000)		(0.0000)
Number of searches	0.0061***	0.0051***	0.0105***	0.0097***
	(0.0003)	(0.0004)	(0.0003)	(0.0004)
R^2	0.11	0.10	0.26	0.26
Number of Observations	2,988	2,315	3,231	2,435

Standard errors in parentheses.

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

[a] Operationalized on a logarithmic scale.

Table 6: Effects of Fatigue Proxies on Search Gap Incidence

	Dependent variable:			
	Search gap indicator			
	(i)	(ii)	(iii)	(iv)
Subcategory	"Shirts, top and blouses"		"Shoes"	
Cumulative number of searched websites	0.3213***		0.0832***	
	(0.0277)		(0.0117)	
Total time spent searching		0.0151***		0.0057***
		(0.0018)		(0.0011)
Spell and Website FE	Yes	Yes	Yes	Yes
Control: Session with Transactions	Yes	Yes	Yes	Yes
R^2	0.46	0.30	0.34	0.32
Number of Observations	7,102	7,102	15,554	15,554

Standard errors in parentheses, clustered at the spell level.

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

Table 7: Monte Carlo Simulation Results

	(i)	(ii)		(iii)	
		Our model		*Weitzman*	
	True values	Estimates	Std. Dev.	Estimates	Std. Dev.
Utility					
Website 1	-1	-0.91	(0.03)	-0.93	(0.02)
Website 2	-0.5	-0.48	(0.03)	-0.48	(0.03)
Website 3	-0.3	-0.30	(0.04)	-0.29	(0.02)
Outside option	0.5	0.54	(0.03)	0.23	(0.04)
Search cost (exp)					
Baseline	-3.5	-3.53	(0.12)	-2.18	(0.04)
Fatigue	-3	-2.67	(0.03)		
Log-likelihood		-19,299		-17,939	
Number of Observations		25,000		25,000	

Table 8: Estimation results

	(i) "Shirts, tops, & blouses"				(ii) "Shoes"		
	Accounting for search gaps		Ignoring search gaps		Accounting for search gaps		Ignoring search gaps
	Our model	*Our model*	*Weitzman*		*Our model*	*Our model*	*Weitzman*
Utility				*Utility*			
aboutyou.com	-1.4368***	-1.3084***	-1.4061***	adidas.com	-1.3895***	-1.3979***	-1.2570***
	(0.0440)	(0.0337)	(0.0463)		(0.0507)	(0.0259)	(0.0423)
c-and-a.com	-0.8623***	-0.7662***	-0.8250***	debijenkorf.nl	-1.9646***	-2.1189***	-1.8366***
	(0.0372)	(0.0288)	(0.0397)		(0.0625)	(0.0467)	(0.0513)
debijenkorf.nl	-1.0226***	-1.3782***	-1.0022***	nelson.nl	-1.9438***	-1.8941***	-1.7978***
	(0.0375)	(0.0406)	(0.0408)		(0.0578)	(0.0416)	(0.0532)
esprit.nl	-1.7118***	-1.7419***	-1.6847***	nike.com	-1.5095***	-1.5162***	-1.3740***
	(0.0496)	(0.0459)	(0.0517)		(0.0540)	(0.0285)	(0.0433)
hm.com	-1.3601***	-1.3584***	-1.3119***	omoda.nl	-1.8658***	-1.8540***	-1.7174***
	(0.0412)	(0.0373)	(0.0444)		(0.0581)	(0.0373)	(0.0491)
jbfo.nl	-2.6664***	-2.4721***	-2.7581***	schuurman-shoenen.nl	-1.1034***	-1.0502***	-0.9748***
	(0.1448)	(0.1261)	(0.1455)		(0.0500)	(0.0211)	(0.0373)
msmode.nl	-2.0188***	-1.9307***	-2.0093***	spartoo.nl	-1.4858***	-1.4437***	-1.3663***
	(0.0653)	(0.0573)	(0.0638)		(0.0514)	(0.0269)	(0.0419)
peterhahn.nl	-2.1520***	-2.0599***	-2.1132***	vanharen.nl	-1.2167***	-1.2168***	-1.0641***
	(0.0816)	(0.0708)	(0.0776)		(0.0466)	(0.0242)	(0.0392)
your-look-for-less.nl	-1.7526***	-1.6579***	-1.7228***	zalando.nl	-0.8434***	-0.9784***	-0.6502***
	(0.0501)	(0.0416)	(0.0512)		(0.0442)	(0.0218)	(0.0344)
zalando.nl	-1.0483***	-1.1357***	-0.9856***	ziengs.nl	-2.0965 ***	-2.1204***	-1.9532***
	(0.0381)	(0.0325)	(0.0410)		(0.0722)	(0.0546)	(0.0578)
Number of previous		0.1407***		Number of previous		0.1849 ***	
website visits		(0.0109)		website visits		(0.0168)	
Visit to a		1.3058***		Visit to a		0.9307***	
price discount page		(0.0411)		price discount page		(0.0442)	
Outside option	1.5823***	1.4050***	1.9583***	Outside option	1.3288***	1.2714***	2.1653***
	(0.0304)	(0.0194)	(0.0512)		(0.0334)	(0.0185)	(0.0558)
Search cost (exp)				*Search cost (exp)*			
Baseline	-5.0740***	-4.8991***	-3.1639***	Baseline	-5.5267***	-5.6355***	-4.2142***
	(0.3024)	(0.2507)	(0.1066)		(0.2950)	(0.2290)	(0.1354)
Fatigue constant	-2.4436***	-2.4793***		Fatigue constant	-2.6379***	-2.6422***	
	(0.0243)	(0.0164)			(0.0395)	(0.0207)	
Fatigue age (≥ 50)		0.3337***		Fatigue age (≥ 50)		0.1131***	
		(0.0268)				(0.0265)	
Number of Observations	27,924	27,600	27,924		27,756	27,264	27,756
LL	-9,356	-8,643	-8,642		-12,179	-11,671	-10,985

Standard errors in parentheses.

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

Table 9: RMSE Comparison

	(i) "Shirts, tops, & blouses"		(ii) "Shoes"	
	Our model	*Weitzman*	*Our model*	*Weitzman*
Number of search gaps	242	763	44	1368
Website search shares	39	52	76	80
Website market shares	71	45	105	49

Table 10: Effects of Fatigue and Search Cost Reductions

	(i) Fatigue reduction by 50%	(ii) Fatigue reduction by 90%	(iii) Search cost reduction by 50%
Subcategory 1: "Shirts, tops, & blouses"			
Percent change			
Number of searches	1.33	18.54	2.02
Number of purchases	0.53	4.51	0.37
Number of search gaps	-11.53	-64.37	6.62
Subcategory 2: "Shoes"			
Percent change			
Number of searches	4.12	28.18	1.32
Number of purchases	1.20	4.99	0.18
Number of search gaps	-22.46	-66.97	3.44

Table 11: Effects of No Breaks

	(i) "Shirts, tops, & blouses"	(ii) "Shoes"
Percent change		
Number of searches	-21.33	-23.11
Number of purchases	-7.65	-5.96

Web Appendix A: Construction of the Final Data Sample

The raw data contain information on the user (including demographics such as age and gender), session, and time of the click, as well as the website name and the entire URL address of the visited website. Furthermore, GfK coded the transaction funnel, identifying a website visit, a product view, a basket addition, a checkout, or an order confirmation.

Data Augmentation: Scraping

Using the full URLs, we attempted to scrape the top 50 fashion websites (ranked by the number of clicks) and were successful in scraping 44 of them. We were able to scrape information from the most popular websites, such as Zalando and H&M, obtaining product information from the top websites accounting for more than 57% of all fashion clicks. This data collection stage occurred within one month of the last observation day in our sample to prevent changes in the web pages. The information we gathered by scraping contains: price, price promotion (if any), page title, brand name, product name, and, if available, product color, reviews, star rating, number of photos, product description, shipping information, speed score of the website (page loading speed), as well as word counts, sentiment on the page, and reading ease.

Data Classification: List, Product, and Other Pages

Although GfK coded the transaction funnel, we performed an additional step in classifying clicks into list, product, or other pages, to ensure that we correctly identify the product (if any) that was purchased. An example of a list page is "https://www.adidas.com/us/women-originals-shoes", where consumers can see a list of shoes along with a photo, the product name and its price. If a consumer clicks on a product in this list, she navigates to that product's page. An example of such a product page is "https://www.adidas.com/us/adidas-sleek-shoes/EE4723.html" for a consumer who clicked on the product "adidas sleek shoes" on the list page. The product page contains more detailed information about the product, such as a product description, additional photos, reviews, etc. Of the 428,651 fashion clicks in our data, 172,536 clicks are on list pages and 93,463 clicks are on product pages. We labeled the remaining clicks as "others" to represent clicks to the homepage of a website, account pages, or any transaction-related pages, such as the cart page.

To categorize clicks into either product, list, or other pages, we performed the following steps.

First, during data scraping, we identified list and product pages by examining whether there was any product information available on the page and, if so, how many products were available on the page. Second, we used the following rules:

- 'Other' pages:
  1. Pages labeled as 'Add to Basket', 'Start Checkout' and 'Order Confirmation' by GfK
  2. The homepage of a website such as 'www.zara.com'
- 'Product' page:
  1. Pages labeled as 'Product View' by GfK
  2. Pages from which we can scrape a single product's information
  3. URLs that contains product SKU or product IDs (rules differ for each website)
  4. URLs with specific keywords such as 'product-view' or 'shop-by-item'
- 'List' page:
  1. Pages from which we can scrape information on multiple products
  2. Pages that display search results, e.g., 'https://www.adidas.com/us/search?q=redshoes'
  3. URLs with specific keywords that indicate their function as a list page, such as 'shop-per-categorie', 'page=', or 'category='
  4. URLs with specific keywords that indicate the sorting or filtering function available on the page, such as 'price_max=', 'productsoort', 'pagenumber' or 'filter='

We further categorized URLs labeled as 'others' by manually checking them and hired an RA to independently manually check our categorization.

Data Classification: Product Categories

To identify the searched product category, we used URLs, page titles, and the scraped information (e.g., product description) to search for keywords identifying nine broad categories (as defined on the most popular website in our data, Zalando): accessories, children's, dresses & skirts, jackets & vests, pants & jeans, shirts & tops and blouses, shoes, sweater and underwear. The keywords used to identify the product categories include but are not limited to:

- **accessories**: 'accesso', 'sjaal', 'lippen', 'earr', 'necklace', 'jewelry', 'bracelet', 'bag', 'eastpak', 'hals', 'banden'
  - with exception of: 'brand', 'bracelet', 'braad', 'brax', 'dirk', 'brace', 'overnachtingen', 'aangebrachte'
- **children's**: 'jongens', 'kinder', 'meisjes', 'baby', 'tiener', 'kids', 'boys' 'girls'
- **dresses & skirts**: 'roecke', 'jurken', 'dress', 'jumpsuit', 'jurkje', 'jurk', 'rok'

- **jackets & vests**: 'trench', 'jack', 'fleece', 'blazer', 'mantel', 'coat', 'parka', 'tussenjas', 'winterjas', 'jas'
- **pants & jeans**: 'hosen', 'broek', 'jogger', 'tights', 'shorts', 'sweatpant', 'pants', 'pantalon', 'leggin', 'trouser', 'tregging', 'jegging'
  - with exception of: 'brand', 'bracelet', 'braad', 'brax', 'dirk', 'brace', 'overnachtingen', 'aangebrachte'
- **shirts, tops & blouses**: 'top', 'hemden', 'langarm', 'kurzarm', 'blusen', 'shirt', 'singlet', 'blouse', 'blouson', 'polo-', '-polo', 'polos', 'longsleeve', 'overhemd', 'onderhemd'
  - with exception of: 'topseller', 'topic', 'topbox', 'topgear', 'topdeals', 'topper', 'topman', 'sniztop', 'marc-c-polo', 'topcom', 'topbloemen', 'laptop'
- **shoes**: 'schuhe', 'stiefel', 'schoen', 'shoe', 'sneaker', 'sandal', 'birkenstock', 'fitflop', 'teva', 'footwear', 'e-walk', 'ecco', 'gabor', 'instappe', 'pumps'
  - with exception of domain names that contain the word 'shoe'
- **sweater**: 'parka', 'hoodie', 'poncho', 'westen', 'trui', 'capuchon', 'pullover', 'tuniek', 'vest', 'cardigan', 'sweater', 'jumper'
- **underwear**: 'thong', 'nightwear', 'bra', 'lingerie', 'sleep', 'swim', 'badpak', 'ondergoed', 'underwear', 'panties', 'sock', 'sok', 'bustier', 'push-up', 'boxer', 'badmode', 'bikini', 'tanga', 'tankini'

Data Classification: Activities

GfK classifies clicks into activities such as 'fashion,' 'social networking,' or 'web search.' We used this classification as well as the following rules to further identify the type of online activity the consumer engaged in. This process resulted in ten categories.

1. Fashion:
   - GfK's classification as 'Fashion'
2. Search engine:
   - GfK's classification as 'Web Search'
   - URLs that contain the keyword 'search' when the website visited is Google, Yahoo or Bing.
   - URLs where the website is 'ask.com'
3. Email
   - GfK's classification as 'Communication'
   - URLs that contain keywords 'mail.google', 'outlook', and 'webmail'.
   - URLs that contain the keyword 'mail' when the visited website is Google, Yahoo or Bing
   - URLs that contain the keyword 'messenger' when the visited website is Yahoo
4. Social Networking

- GfK's classification as 'Social Networking'
- The visited website is one of the 5 major social media platforms: facebook, pinterest, twitter, instagram, linkedin

5. Banking
   - GfK's classification as 'Money Management'
   - The visited website is or contains 'rabobank.nl', 'abnamro.nl', 'bank', 'achmea' or 'vanlanschot'
6. Cashback
   - The visited website is one of: 'geldrace.nl', 'geldkoffer.info', 'geldwolf.info', 'zinngeld.nl', 'mailbeurs.nl', 'extraeuro.nl', 'centmail.nl', 'cashhier.nl', 'spaar4cash.nl', 'snelverdienen.nl', 'ipay.nl', 'spaaractief.nl', 'nucash.nl', 'myflavours.nl', 'directverdiend.nl', 'dieselmail.nl', 'spaar4cash.nl', 'dutcheuro.nl', 'extraeuro.nl', 'cashparadijs.nl', 'sneleuro.nl', 'myclics.nl', 'spaar-voor-euries.nl', 'jiggy.nl', 'qlics.nl', 'quidco'
7. Surveys
   - The visited website is one of: 'gfk.com', 'ssisurveys.com', 'focusvision.com', 'opinionbar.com', 'globaltestmarket.com'
8. Media
   - GfK's classification as 'Media Broadcasting' or 'Media On-Demand'
   - URLs that contain the keyword 'tvgids'
9. Google exclude
   - URLs from 'google.com' that are not classified as search engine or email related (this includes Google Drive, Maps, etc.)
10. Gaming
    - GfK's classification as 'Gaming'
    - URLs containing the keywords: 'casino', 'game', 'unibet', 'nederlandseloterij.nl'

Data Cleaning: Removing Non-search Activity

The raw data contains 9,531,448 observations. To obtain the final data set, we removed observations in the following cases:

- Consumers use a web browser to open local files on their computers rather than browse the Internet
- A new tab is opened but no webpage is visited on that tab
- Consumers open web browsers' extensions
- Any URL that does not contain 'ttp'

- Duplicates at the session-time level: the same URL is clicked more than once at the same time or two different URLs are clicked at the same time. In both cases, we only kept a record of the first click.
- Spells during which sessions overlap in time (one instance)
- Spells with a transaction but no clicks on product pages observed (in these rare cases, websites likely offer the option of adding a product to the cart directly from the list page)
- Spells with a transaction and observed product clicks but no product added to the cart
- Spells that end within the first week of our observation period, i.e., before February 23rd, 2018, since it is likely that these observations are left truncated.

These changes resulted in a data set with 7,877,551 observations. In addition, among the fashion clicks (437,659 observations), we dropped sessions and their corresponding spells that only contained clicks that were unrelated to product search activity, such as clicks to log in or out of an account, to track a shipment, to find a store location, to access customer service, to manage a subscription or to create a password. Of the original 437,659 observations, we were left with 428,651 observations.

Estimation Samples

We constructed our estimation samples as follows. We focused on the two most commonly purchased product subcategories in our data: "shirts, tops, & blouses" and "shoes." We removed search revisits (i.e., revisits to the same website beyond the first visit that are unrelated to the actions required to make a purchase, such as logging in into an account) from the sample (approximately 30% of observations). To address concerns about right truncation, we removed spells that did not end in a transaction and that had a search session within the last two days of our observation period. We focused on spells with at most one transaction (more than 99.3% of spells) and we removed clicks that might have occurred after the consumer saw an ad on social media, email, newsletter or through retargeting (fewer than 20% of spells).[40] For each subcategory, we determined the top 10 most searched websites (accounting for approximately 65% of clicks in each subcategory), for which we estimate website intercepts. All other websites were grouped together into a composite website which we call "Other."

The resulting estimation samples have 27,924 observations and 2,327 spells in the "shirts, tops, & blouses" subcategory, and 27,756 observations and 2,313 spells in the "shoes" subcategory. Consumers made 309 and 248 transactions in each subcategory, respectively. Further, there are 763 and 1,368 search gaps in each subcategory, respectively, with 586 and 818 spells with at least one search gap.

[40] Note that ads may encourage consumer to start a new search session, but cannot explain why consumers stopped their search in a previous session, and therefore cannot explain search gaps.

Web Appendix B: Estimation Details

Estimation procedure: Our model

The estimation using the logit-smoothed AR simulator involves the following steps:

1. Make $d = \{1, \ldots, D\}$ draws of η_{ij} and ϵ_{ij} for each consumer-website combination and calculate utility u_{ij}^d.
2. Compute $\left[z_j^d,\, z_j^{1d}(t),\, z_j^{2d},\, z_j^{3d}(t)\right]$
3. Calculate the following expressions for each draw d:
   (a) $v_1^d = z_{in}^d - \max_{k=n+1}^{J} z_{ik}^d \qquad \forall n \in \{1, \ldots, J-1\}$
   (b) $v_2^d = z_{i1}^d - u_{i0}^d$
   (c) $v_3^d = \max\left\{z_{in}^{1d}(t_n), z_{in}^{2d}\right\} - \max_{k=0}^{n-1} u_{ik}^d \qquad \forall n \in \{2, \ldots, s\}$
   (d) $v_4^d = \max_{k=0}^{n-1} u_{ik}^d - z_{in}^{3d}(t_n)$ if $t_n = t_{n+1} - 1 \qquad \forall n \in \{2, \ldots, s\}$
   (e) $v_5^d = z_{in}^{3d}(t_n) - \max_{k=0}^{n-1} u_{ik}^d$ if $t_n \neq t_{n+1} - 1 \qquad \forall n \in \{2, \ldots, s\}$
   (f) $v_6^d = z_{im}^{1d}(t_m) - \max_{k=0}^{s} u_{ik}^d \qquad \forall m \in \{s+1, \ldots, J\}$
   (g) $v_7^d = z_{im}^{2d} - \max_{k=0}^{s} u_{ik}^d \qquad \forall m \in \{s+1, \ldots, J\}$
   (h) $v_8^d = u_{ij}^d - \max_{k=0}^{s} u_{ik}^d \qquad \forall j \in \{0, 1, \ldots, s\}$
4. Compute $V^d = \frac{1}{1+M^d}$ for each draw d, where

$$M^d = e^{-v_1^d/\rho_1} + e^{-v_2^d/\rho_2} + (e^{-v_3^d/\rho_3} + e^{-v_4^d/\rho_3}) + (e^{-v_3^d/\rho_4} + e^{-v_5^d/\rho_4}) + (e^{-v_6^d/\rho_5} + e^{-v_7^d/\rho_5}) + e^{-v_8^d/\rho_6}, \tag{B1}$$

   where the terms in parentheses represent the values of searching now, searching later, and of stopping, respectively, and where ρ_k are scaling parameters, chosen using Monte Carlo simulations (Honka 2014; Ursu 2018; Ursu, Wang, and Chintagunta 2020).
5. The average of V^d over the D draws and over consumers and websites gives the simulated likelihood function.

Similar to Ursu, Wang, and Chintagunta 2020, we use different scaling values ρ_k for each for 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: $\rho = [-10, -3, -3, -10, -10, -5]$. Therefore, we estimate our model with the same set of scaling values. We also notice that estimating the model with all scaling parameters set to -3 also recovers the parameters well.[41]

[41] The analysis is available from the authors upon request.

Estimation procedure: Weitzman model

In the Weitzman model consumers search products in order of their reservation utilities and stop searching when the best observed utility so far exceeds the reservation utility of any unsearched option. The estimation procedure using the logit-smoothed AR simulator follows that in Honka and Chintagunta (2017) and Ursu (2018) and involves the following steps:

1. Make $d = \{1, \ldots, D\}$ draws of η_{ij} and ϵ_{ij} for each consumer-website combination and calculate utility u_{ij}^d.
2. Compute z_j^d.
3. Calculate the following expressions for each draw d:
   (a) $v_1^d = z_{in}^d - \max_{k=n+1}^{J} z_{ik}^d \qquad \forall n \in \{1, \ldots, J-1\}$
   (b) $v_2^d = z_{in}^d - \max_{k=0}^{n-1} u_{ik}^d \qquad \forall n \in \{1, \ldots, s\}$
   (c) $v_3^d = \max_{k=0}^{s} u_{ik}^d - z_{im}^d \qquad \forall m \in \{s+1, \ldots, J\}$
   (d) $v_4^d = u_{ij}^d - \max_{k=0}^{s} u_{ik}^d \qquad \forall j \in \{0, 1, \ldots, s\}$
4. Compute $V^d = \frac{1}{1+M^d}$ for each draw d, where

$$M^d = \sum_{k=1}^{4} e^{-v_k^d/\rho_k}, \tag{B2}$$

   where ρ_k are scaling parameters, chosen using Monte Carlo simulations (Honka 2014; Ursu 2018; Ursu, Wang, and Chintagunta 2020).
5. The average of V^d over the D draws and over consumers and websites gives the simulated likelihood function.

To ensure consistency across the estimation of the two models, we use the same scaling parameters, adjusted for the fact that, in the Weitzman model, the likelihood function is made up of only four components (rather than eight as in our model). The set of scaling parameters is given by $\rho = [-10, -3, -3, -5]$.

Web Appendix C: Calculating Reservation Utilities

An advantage of our proposed method lies in its ease of estimation due to its similarity to the Weitzman (1979) model: consumers search in the decreasing order of reservation utilities z_j and also make search and purchase decisions based on threshold values of the best observed alternative so far. The main difference consists of computing the values of $\left[z_j^1(t), z_j^2, z_j^3(t)\right]$ in addition to that of z_j. We start by describing the method used in Kim, Albuquerque, and Bronnenberg 2010 to compute z_j. Then we describe our method to compute $\left[z_j^1(\cdot), z_j^2, z_j^3(\cdot)\right]$.

Recall that $W_j(y) = yF_j(y) + \int_y^\infty u dF_j(u)$ and from Equation (8) that the reservation utility z_j is the solution to $c_{j0} = W_j\left(z_j\right) - z_j$. From Kim, Albuquerque, and Bronnenberg (2010), we know that, under the assumption that ϵ_j is standard normally distributed,

$$\begin{aligned} B\left(m_j\right) &= W_j\left(z_j\right) - z_j \\ &= \phi\left(m_j\right) + m_j\Phi\left(m_j\right) - m_j \end{aligned}$$

with $m_j = z_j - \mu_j$. Given that a unique solution to $c_{j0} = B\left(m_j\right)$ exists (see Weitzman 1979 or our discussion in Section 5.4), one can invert the relation, solve for m_j and then compute z_j from the relation $z_j = m_j + \mu_j$. Following prior work, we create a look-up table relating c_{j0} to m_j according to function $B\left(m_j\right)$, which we can use to solve for z_j for any search cost value.

To compute $z_j^1(t)$, we use a similar method. Recall from Equation (9) that $z_j^1(t)$ is the solution to

$$c_{j0} + (\alpha - \delta)t = W_j\left(z_j^1(t)\right) - z_j^1(t). \tag{C1}$$

Since the additional term affecting search costs is constant in j, we can similarly create a look-up table relating $c_{j0} + (\alpha - \delta)t$ to $m_j^1(t)$ according to the function $B\left(m_j^1(t)\right)$ for the observed value of t (same function $B(\cdot)$ as above), and then solve for $z_j^1(t)$ from $z_j^1(t) = m_j^1(t) + \mu_j$.

To compute $\left[z_j^2, z_j^3(t)\right]$, we use a different method. We start by noting that if $W_j(z) - z = \phi(m) + m\Phi(m) - m$, then it follows that $W_j(z) = \phi(m) + m\Phi(m) + \mu$ for $m = z - \mu$. Next, recall that z_j^2 is the solution to

$$
\begin{aligned}
c_{j0} &= W_j\left(z_j^2\right) - \frac{z_j^2}{\beta} \\
&= \phi\left(z_j^2 - \mu_j\right) + \left(z_j^2 - \mu_j\right)\Phi\left(z_j^2 - \mu_j\right) + \mu_j - \frac{z_j^2}{\beta}. \qquad \text{(C2)}
\end{aligned}
$$

To solve for z_j^2, we create a look-up table relating c_{j0} and μ_j to values of z_j^2 directly by (numerically) solving the stated equation for all relevant values of c_{j0} and μ_j.

To solve for $z_j^3(t)$, we use a similar approach. From Equation (11), we know $z_j^3(t)$ is the solution to

$$
c_{j0} = W_j\left(z_j^3(t)\right) - \frac{(\alpha-\delta)\,t}{1-\beta}, \qquad \text{(C3)}
$$

which can be rewritten as

$$
c_{j0} + \frac{(\alpha-\delta)\,t}{1-\beta} = \phi\left(z_j^3(t) - \mu_j\right) + \left(z_j^3(t) - \mu_j\right)\Phi\left(z_j^3(t) - \mu_j\right) + \mu_j. \qquad \text{(C4)}
$$

Using a look-up table relating $c_{j0} + \frac{(\alpha-\delta)t}{1-\beta}$ and μ_j to values of $z_j^3(t)$, after (numerically) solving the stated equation, we can compute $z_j^3(t)$.

Web Appendix D: Robustness Checks

Weitzman Model with Increasing Search Cost

Our model makes two changes to the Weitzman framework: (i) allows for search gaps; and (ii) models the effect of fatigue on search costs. To better isolate the effect of each change on parameter estimates, we also estimate a variation of the Weitzman model with increasing search costs (due to fatigue) but without search gaps, i.e., we only make one change to the Weitzman framework. Technically, this involves removing the option to search after the break from Equation 3, but continuing to assume that fatigue affects search costs. To the best of our knowledge, this variation of the Weitzman model has not been studied by previous literature. As in our problem, there is no known optimal search rule. However, the solution we developed for our model can be used to derive an optimal search rule for this variation of the Weitzman model. Using the argument made for Theorem 2 (Case 1), under Condition 1, the consumer searches products in the same order as in the original Weitzman

model, i.e., in decreasing order of their reservation utilities z_j. And, the consumer stops searching when she encounters a product j for which $z_j^1(t)$ (defined as in Equation C1) is smaller than the best option searched so far, and continues searching j otherwise.

Table D-1: Estimation results for Weitzman model with increasing search costs

	(i) "Shirts, tops, & blouses" *Adapted Weitzman*		(ii) "Shoes" *Adapted Weitzman*
Utility		*Utility*	
aboutyou.com	-1.3920***	adidas.com	-1.2832***
	(0.0465)		(0.0408)
c-and-a.com	-0.8181***	debijenkorf.nl	-1.8562***
	(0.0400)		(0.0488)
debijenkorf.nl	-0.9952***	nelson.nl	-1.8205***
	(0.0410)		(0.0493)
esprit.nl	-1.6627***	nike.com	-1.3993***
	(0.0519)		(0.0415)
hm.com	-1.3008***	omoda.nl	-1.7360***
	(0.0447)		(0.0467)
jbfo.nl	-2.7536***	schuurman-shoenen.nl	-0.9964***
	(0.1486)		(0.0360)
msmode.nl	-2.0303***	spartoo.nl	-1.3912***
	(0.0643)		(0.0402)
peterhahn.nl	-2.1501***	vanharen.nl	-1.0869***
	(0.0787)		(0.0378)
your-look-for-less.nl	-1.7080***	zalando.nl	-0.6744***
	(0.0514)		(0.0329)
zalando.nl	-0.9790***	ziengs.nl	-1.9833***
	(0.0412)		(0.0556)
Outside option	1.9572***	Outside option	2.1079***
	(0.0475)		(0.0481)
Search cost (exp)		*Search cost (exp)*	
Baseline	-3.8829***	Baseline	-4.3036***
	(0.0043)		(0.0823)
Number of previous searches	-3.9986***	Number of previous searches	-7.7467***
	(0.0229)		(1.0988)
Number of Observations	27,924		27,756
LL	-8,644		-10,990

Standard errors in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$

Our Model with $\beta = 0.9$

Table D-2: Estimation results for $\beta = 0.9$

	(i) "Shirts, tops, & blouses" *Our model*		(ii) "Shoes" *Our model*
Utility		*Utility*	
aboutyou.com	-1.4264***	adidas.com	-1.4067***
	(0.0452)		(0.0523)
c-and-a.com	-0.8654***	debijenkorf.nl	-2.0119***
	(0.0393)		(0.0635)
debijenkorf.nl	-1.0263***	nelson.nl	-1.9857***
	(0.0395)		(0.0581)
esprit.nl	-1.7031***	nike.com	-1.5281***
	(0.0516)		(0.0525)
hm.com	-1.3484***	omoda.nl	-1.9050***
	(0.0427)		(0.0579)
jbfo.nl	-2.7816***	schuurman-shoenen.nl	-1.1262***
	(0.1582)		(0.0493)
msmode.nl	-2.0354***	spartoo.nl	-1.4976***
	(0.0656)		(0.0502)
peterhahn.nl	-2.1891***	vanharen.nl	-1.2346***
	(0.0820)		(0.0484)
your-look-for-less.nl	-1.7456***	zalando.nl	-0.8707***
	(0.0502)		(0.0428)
zalando.nl	-1.0445***	ziengs.nl	-2.1667***
	(0.0400)		(0.0711)
Outside option	1.5634***	Outside option	1.2679***
	(0.0344)		(0.0366)
Search cost (exp)		*Search cost (exp)*	
Baseline	-4.9558***	Baseline	-5.0183***
	(0.3286)		(0.2267)
Fatigue constant	-1.7889***	Fatigue constant	-2.0332***
	(0.0342)		(0.0512)
Number of Observations	27,924		27,756
LL	-9,333		-12,157

Standard errors in parentheses.
* $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$