DISCUSSION PAPER SERIES

DP16193

Online Advertising as Passive Search

Raluca Ursu, Andrey Simonov and Eunkyung An

INDUSTRIAL ORGANIZATION

CEPR

ISSN 0265-8003

Online Advertising as Passive Search

Raluca Ursu, Andrey Simonov and Eunkyung An

Discussion Paper DP16193
Published 26 May 2021
Submitted 26 May 2021

Centre for Economic Policy Research
33 Great Sutton Street, London EC1V 0DX, UK
Tel: +44 (0)20 7183 8801
www.cepr.org

Online Advertising as Passive Search

Abstract

Standard search models assume that consumers actively decide on the order, identity, and number of products they search. We document that online, a large fraction of searches happen in a more passive manner, with consumers merely reacting to online advertisements that do not allow them to choose the timing or the identity of products to which they will be exposed. Using a clickstream panel data set capturing full URL addresses of websites consumers visit, we show how to detect whether a click is ad-initiated. We then document that ad-initiated clicks account for more than half of all website arrivals, are more concentrated early on in the consumer search process, and lead to less in-depth searches and fewer transactions, consistent with the passive nature of these searches. To account for these systematic differences between active and passive searches, we propose and estimate a simple model that accommodates both types of searches, and describe the estimation bias arising in models that incorrectly treat all searches as active. Finally, we use our model's estimates to describe consumer substitution patterns across websites under different advertising scenarios.

JEL Classification: D83, L81, M31, M37

Keywords: sequential search, advertising, Online Browsing, Fashion Industry

Raluca Ursu - rmu208@stern.nyu.edu
NYU Stern

Andrey Simonov - as5443@gsb.columbia.edu
Columbia Business School and CEPR

Eunkyung An - ean@stern.nyu.edu
NYU Stern

Acknowledgements
Authors contributed equally and names are listed in reverse alphabetical order. We are thankful for comments from Yuxin Chen, Tulin Erdem, Anindya Ghose, Kinshuk Jerath, Eitan Muller, Oded Netzer, Stephan Seiler, Olivier Toubia, and attendees of the IIOC 2021 conference. We would also like to thank GfK and Siham El Kihal for helping us obtain the data. The usual disclaimer applies.

Online Advertising as Passive Search*

Raluca M. Ursu[†] Andrey Simonov[‡] Eunkyung An[§]

May 6, 2021

Abstract

Standard search models assume that consumers actively decide on the order, identity, and number of products they search. We document that online, a large fraction of searches happen in a more passive manner, with consumers merely reacting to online advertisements that do not allow them to choose the timing or the identity of products to which they will be exposed. Using a clickstream panel data set capturing full URL addresses of websites consumers visit, we show how to detect whether a click is ad-initiated. We then document that ad-initiated clicks account for more than half of all website arrivals, are more concentrated early on in the consumer search process, and lead to less in-depth searches and fewer transactions, consistent with the passive nature of these searches. To account for these systematic differences between active and passive searches, we propose and estimate a simple model that accommodates both types of searches, and describe the estimation bias arising in models that incorrectly treat all searches as active. Finally, we use our model's estimates to describe consumer substitution patterns across websites under different advertising scenarios.

Keywords: Sequential Search, Advertising, Online Browsing, Fashion Industry

JEL Classification: D83, L81, M31, M37

---

*Authors contributed equally and names are listed in reverse alphabetical order. We are thankful for comments from Yuxin Chen, Tulin Erdem, Anindya Ghose, Kinshuk Jerath, Eitan Muller, Oded Netzer, Stephan Seiler, Olivier Toubia, and attendees of the IIOC 2021 conference. We would also like to thank GfK and Siham El Kihal for helping us obtain the data. The usual disclaimer applies.

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

[‡]Columbia Business School and CEPR, as5443@gsb.columbia.edu.

[§]Stern School of Business, New York University, ean@stern.nyu.edu.

1 Introduction

In recent years, the ready availability of consumer clickstream data has generated an unprecedented level of interest in the study of online consumer search decisions (e.g., Kim et al., 2010; De los Santos et al., 2012; Koulayev, 2014; Bronnenberg et al., 2016; Chen and Yao, 2017; Ursu, 2018). These clickstream data reveal sequences of products searched, which – together with models of sequential search (e.g., Weitzman, 1979) – help researchers recover preference and search cost estimates, and thereby inform marketing and economic decisions. A key assumption made in these sequential search models is that consumers *actively* decide on the order, identity, and number of all products they search.

However, a large fraction of online searches are initiated by advertisements – U.S. companies spend more than half of their total advertising budgets ($129 billion in 2019) on online marketing strategies, such as paid search, email marketing, or display ads.[1] Ads do not allow consumers to choose the timing or the identity of products to which they will be exposed, and consumers frequently encounter ads when engaged in activities that are unrelated to product search (e.g. when browsing social media websites or reading the news).[2] Thus, when search is ad-initiated, consumers may search in a more *passive* manner than assumed in standard search models – i.e. they choose how to react to information to which they were exposed (e.g. whether to click), but do not optimally choose what information to see or in what order to see it.[3]

In this paper, we study the role of online advertising in consumers' information search process by distinguishing between active and ad-initiated searches. We employ a detailed clickstream data set capturing all web traffic (8 million clicks) of a panel of 4,600 consumers in the Netherlands at the level of the exact URL address of a website visited. These data contain clicks in our focal category – fashion – as well as all other online activities consumers performed, such as checking email, visiting Facebook, or reading the news. A special feature of our data is the granularity of the URL addresses captured – they contain not only information on the webpage accessed (e.g. www.nike.com), but also information on how the consumer landed on that webpage. In particular, these URLs contain specific keywords

[1] For more information, see emarketer.com/us-digital-ad-spending.

[2] In 2018, 78% of Facebook users discovered new products while browsing the site, and 55% of Americans bought products online after a social media ad. For more details, see adespresso.com/facebook-statistics and kleinerperkins.com/internet-trends-report-2018.

[3] We share the terminology of "active" and "passive" search with Renault (2016) and Ghose and Todri-Adamopoulos (2016), defining active search as the effortful action to seek out product information optimally, and passive search as the reaction to information to which one is exposed. In both cases consumers choose whether to search a product. The difference is that, in contrast to active search, under passive search consumers do not choose which product to search next (i.e. the optimal search order). Similar ideas appear in Honka et al. (2017) and Morozov (2020).

that identify the advertiser and the medium of advertising (e.g. email, display, social media) in cases when the consumer landed on a webpage through an ad. Using these data, we develop a method for detecting ad-initiated clicks, and describe and separate them from searches occurring organically. The proposed method enables our investigation and distinguishes our paper from prior work on consumer search that does not have access to such granular data.[4]

We then document the volume and describe the nature of ad-initiated searches. Product searches initiated by ads are extensive – 15% of all clicks and 53% of all website arrivals are a result of clicks on ads, with some variation across product categories and websites. Consumers are more likely to search through ads early on in their search process – the probability that a search is ad-initiated is 22% in the first decile of search and only 7% in the last decile. Furthermore, ad-initiated clicks lead to lower quality website searches compared to organic clicks – these website visits involve searches of fewer and more expensive products, are shorter, and are less likely to result in a purchase. Finally, most ad-initiated searches occur when consumers are engaged in online activities that are unrelated to shopping (e.g. when checking email, visiting social media websites, or reading the news).

These patterns further indicate that ad-initiated searches do not align with the active search behavior assumed by standard search models (e.g., Rothschild, 1974; Weitzman, 1979) and used frequently in empirical applications (Kim et al., 2010, 2017; Koulayev, 2014; Chen and Yao, 2017; Ma, 2016; Honka and Chintagunta, 2017; De los Santos and Koulayev, 2017; Ursu, 2018; Gardete and Hunter, 2020; Ursu et al., 2020a). Such models assume consumers choose which options to search next, optimally searching products in decreasing order of an index (reservation utility) representing their expected utility net of search costs. In contrast, we find that ad-initiated searches occur predominantly early in the search process, but are generally of lower quality. Furthermore, it is unlikely that consumers take the indirect route of checking email or visiting social media websites with the purpose of searching for information on fashion products, since searching fashion websites directly is easier. Thus, their response to ads in such shopping-unrelated settings is inconsistent with the notion of active search, which requires that consumers actively seek out product information. We conclude that ad-initiated searches are more in line with the notion of passive search, according to which consumers react to information to which they are exposed, but do not choose what information to see or when to see it.

[4]Typical clickstream data only reveal the information consumers obtained once on a website (e.g. quality, price), but not how they landed on the website (actively or through an ad) (De los Santos et al., 2012; Koulayev, 2014; Chen and Yao, 2017; Ursu, 2018; Gardete and Hunter, 2020; Ursu et al., 2020a).

We propose a simple model to account for both active and passive search decisions. The model builds on the canonical sequential search model of Weitzman (1979) and combines it with insights from the theoretical framework proposed by Renault (2016), where all search is passive. Consistent with the Weitzman (1979) model, the consumer optimally ranks options she is aware of by their reservation utility and proceeds to search in that order, stopping to make a purchase decision when the best option she revealed while searching exceeds the reservation utility of any unsearched option. In addition, in our model the consumer may be exposed to information about an option she is not aware of through advertising, a mechanism documented by prior work (Terui et al., 2011; Honka et al., 2017).[5] Consistent with Renault (2016) and our data patterns, we model consumers' search in response to ads as passive, i.e. consumers do not have control over the timing or the identity of the products they will be exposed to through ads, but only choose how to react to them (whether to search those products).[6] More precisely, in our main specification (*AP-strong model*), we model consumers as merely choosing whether to click on the ad they were exposed to, comparing it with the best option observed so far. We also examine other variations of our model, including a version where consumers compare ads not only to the best option observed so far, but also to other unsearched options (*AP-weak model*), and a version where the advertised products are searched actively (using Weitzman optimal search rules), but may have different search costs.

In our empirical application, we demonstrate the better fit of our proposed models of active and passive search over models that treat all ad-initiated search as active. We model consumers as searching across websites in the four largest fashion subcategories ("shirts, tops, and blouses", "shoes", "pants and jeans", and "underwear"). We find that across all subcategories, our main model of active and passive search (AP-strong) has the best fit, followed by the second variation of the model (AP-weak) and the Weitzman model with different search costs for advertised options. The standard Weitzman model where all searches are treated as active and ads do not affect search costs leads to the worst data

[5] Prior work has documented that the primary mechanism through which advertising affects the consumer search process is awareness (Goeree, 2008; Terui et al., 2011; Honka et al., 2017; Tsai and Honka, 2018). More broadly, our specification is consistent with the literature documenting the informative effect of advertising (Ackerberg, 2001, 2003; Abhishek et al., 2012; Blake et al., 2015; Sahni and Zhang, 2020).

[6] An important aspect of online advertising is that ads can be targeted to a specific consumer, based on, for example, their previous search history. However, as long as consumers are not strategic in searching certain products in order to receive advertising messages from specific brands at specific times, then they will not have full control over the ads they will see – consistent with passive search – behavior which our model is meant to capture. If instead consumers were strategic and could affect the probability or the timing of ads, then it might be more reasonable to treat these searches as active searches as well. We allow for this possibility by estimating the Weitzman model of active search on our data. More details can be found in Section 5, where we also discuss our approach to dealing with firm's strategic advertising decisions.

fit. This result highlights the different nature of ad-initiated searches and the importance of accounting for the role of advertising when modeling search decisions. Further, we find that treating all searches as active leads to biased estimates – consumer preferences for websites that advertise frequently are overestimated by the Weitzman model by 18% compared to our main model. This bias occurs because most ad-initiated searches happen early on in the search process, leading the Weitzman model to incorrectly assume that these options have high reservation utilities. In contrast, our model predicts that advertised websites will be clicked more early in the search process not because of their high reservation utilities, but because the consumer has not yet searched options with high enough utility.

To understand how advertising affects competition between websites, we use our model's estimates to examine consumer substitution patterns under different advertising scenarios. For this, we simulate consumer searches and choices when some or all advertising is shut down. Consistent with our model, we assume that consumers are not aware of advertising websites if ads are shut down, absent the crowd-out behavior described in Blake et al. (2015) and Simonov et al. (2018). We start with the case when all advertising is shut down. We find that with no advertising, we would observe 20-36% fewer searches and 11-21% fewer transactions. This decrease is primarily driven by a loss of approximately 60% in searches and purchases experienced by websites that advertise frequently, with consumers substituting to searching and buying from other websites, some of which benefit from the change. We then investigate two additional scenarios, one with no advertising for websites that advertise frequently and another with no advertising for websites with high conversion rates, and find qualitatively similar effects, albeit smaller in magnitude.

This paper brings together the advertising and consumer search literatures by studying how ad-initiated clicks enter the consumer search process. Our results document the important role of ad-initiated searches – they represent the majority of website visits, occur predominantly early in the search process, but are unlikely to lead to a transaction. These results can help managers account for the different nature of ad-initiated and active clicks. As we have shown, incorrectly assuming that ad-initiated clicks are a result of an active search process may lead to biased parameter estimates, affecting managers' advertising decisions. For example, assuming a consumer has actively searched a product on Nike.com, rather than passively reacted to a Nike ad – even if it is for the same product – implies wrongly assuming the consumer expects Nike's product offerings to dominate those of other brands, inflating consumer brand preferences.

The rest of the paper is organized as follows. Section 2 discusses the related literature. We introduce our data in Section 3. Section 4 provides descriptive results on the nature of ad-initiated searches. Sections 5, 6, and 7 introduce our model, estimation strategy, and results, respectively. Section 8 discusses managerial implications, and Section 9 concludes.

2 Related Literature

2.1 Consumer Search

Our paper relates and contributes to the literature on consumer search. Both theoretical (e.g., Stigler, 1961; Rothschild, 1974; Weitzman, 1979; Wolinsky, 1986; Anderson and Renault, 1999; Branco et al., 2012, 2016; Ke et al., 2016; Dukes and Liu, 2016; Ke and Villas-Boas, 2019) and empirical (e.g., Hong and Shum, 2006; Moraga-González and Wildenbeest, 2008; Ratchford, 2008; Kim et al., 2010, 2017; De los Santos et al., 2012; Seiler, 2013; Honka, 2014; Koulayev, 2014; Moraga-González et al., 2015; Bronnenberg et al., 2016; Chen and Yao, 2017; Ma, 2016; Honka and Chintagunta, 2017; De los Santos and Koulayev, 2017; Ursu, 2018; Gardete and Hunter, 2020; Ursu et al., 2020a) branches of this literature consider consumers who actively decide what products to search, in what order to search them, and whether to purchase the best option found. In these models, consumers exert costly effort to seek out product information, and firms affect this search process only indirectly – for instance, through prices, product features, or product recommedations. Our first contribution to this literature consists of documenting that a large fraction of online consumer searches happen through ads – a channel that does not necessarily align with the assumed active nature of consumer search. We additionally contribute to this literature by developing a method to detect ad-initiated searches from clickstream data and by proposing a model where consumers search both actively and passively.

Our data patterns suggest that ad-initiated searches are unlikely to result from an active search process (see Section 4 for more details). To account for these searches, we build on earlier work that describes consumers' passive information acquisition processes, occurring for example through personal sources (e.g. friends, relatives, neighbors), unsponsored sources (customer reports), as well as marketing dominated sources, such as TV, newspaper, or radio ads (Katona and Mueller, 1955; Bennett and Mandell, 1969; Newman and Staelin, 1972; Newman and Lockeman, 1975; Beales et al., 1981; Duncan and Olshavsky, 1982; Furse et al., 1984; Beatty and Smith, 1987; Shim and Drake, 1989).

This literature also provides empirical observations on passive search – for instance, Beales et al. (1981) explain that passive search can lead consumers to gather different types of information (e.g. about the existence of a product, rather than about prices or other features) – but does not provide a theoretical formalization of a passive search process. To the best of our knowledge, Renault (2016) is the only paper that proposes a model of passive search, describing consumers who decide whether to click on an ad to obtain additional information about a product or whether to wait for another ad. In the model of Renault (2016), all search is passive; we make a contribution by combining it with the Weitzman (1979) model of active search, developing a model of joint active and passive search decisions. We then estimate our model and show that it outperforms one that treats all searches as active.

Our paper also relates to the recent work of Gossner et al. (2020), which studies the role of attention manipulation in the information gathering process of a decision-maker. The authors develop a theoretical model in which techniques such as advertising focus a consumer's attention on one product, thereby affecting the order in which she processes information about other options and when she stops to make a purchase decision. This model predicts that attention accelerates a decision, as well as increases the probability of a consumer choosing the focal product. In contrast, our data show that ads are infrequently purchased and that consumers clicking on more ads generally have longer searches.

Finally, we note that our paper uses the same data as Ursu et al. (2020b), but studies a different question. Namely, in Ursu et al. (2020b) the focus is on understanding why consumers stop and restart their search across sessions. The authors propose that one mechanism affecting this decision is fatigue. Advertising may also explain why consumers restart their search, but it cannot explain why they frequently stop searching. Nevertheless, to be conservative, the authors focus on searches without clicks on the main advertising types. In contrast, understanding the broad role of online advertising is the focus of our paper.

2.2 Advertising and Consumer Search

Our paper also contributes to prior work on advertising and consumer search. On the theoretical side, papers in this literature consider models where either the only source of information consumers have access to arrives through advertising (Iyer et al., 2005), or where consumers can search in a second stage after receiving ads in a first stage (Butters, 1978; Robert and Stahl, 1993; Anderson and Renault, 2006, 2013; Mayzlin and Shin, 2011; De Corniere, 2016; Burguet and Petrikaite, 2017). Such two-stage models implicitly consider both the passive (first stage) and the active (second stage) nature of search. We contribute to this literature by proposing a model where information acquisition through both passive and active searches can happen throughout the entire decision-making process.

In our model, advertising makes consumers aware of new products, consistent with prior work showing that the primary mechanism through which advertising affects the consumer search and choice process is awareness (e.g., Goeree, 2008; Terui et al., 2011; Honka et al., 2017; Tsai and Honka, 2018). Broadly, our paper fits into the literature documenting the informative effects of advertising (Ackerberg, 2001, 2003; Abhishek et al., 2012; Blake et al., 2015; Sahni and Zhang, 2020). The closest paper to ours is Honka et al. (2017), showing that advertising affects awareness, and that consumers engage in (active) search in the consideration stage (second stage). Unlike Honka et al. (2017), our model allows consumers to search actively and passively throughout the decision-making process.

On the empirical side, our paper relates to the rich literature on advertising effects on search (Yang and Ghose, 2010; Rutz et al., 2011; Goldfarb and Tucker, 2011; Rutz and Bucklin, 2012; Narayanan and Kalyanam, 2015; Jeziorski and Moorthy, 2017; Golden and Horton, 2020; Fong, 2017; Rao and Simonov, 2018; Simonov et al., 2018; Du et al., 2019; Joo et al., 2013, 2016; Ghose and Todri-Adamopoulos, 2016; Sahni and Zhang, 2020; Simonov and Hill, 2021). Similar to the theoretical literature we discussed above, in these papers, search occurs only in a second stage after consumers were first exposed to advertising. In contrast, we consider the interplay between active and passive searches throughout the search process, and we develop a structural model of consumer search in the presence of advertising.

Our paper is also related to prior work on advertising attribution models (Abhishek et al., 2012; Li and Kannan, 2014; Kireyev et al., 2016; Chan et al., 2011). These papers propose methods (frequently Hidden Markov Models) to identify the impact of ads at different stages in the consumer conversion funnel, in order to measure the contribution of ads to the final purchase decision. For example, Abhishek et al. (2012) shows that display ads move consumers from a disengaged state to an awareness state, but not further towards a consideration state. Similarly, we model the effect of advertising on awareness. In contrast to this work, we focus on understanding the interaction of advertising and active search. We are able to do this by employing a rich data set of search across brands, capturing the entire browsing behavior of consumers.

3 Data

3.1 Data Description

In this section, we provide an overview of our data. The data were collected by GfK ("Growth from Knowledge"), the largest German market research company. Our data contain the complete PC browsing histories of an online panel of representative consumers from the Netherlands over the time period February 15, 2018 to May 1, 2018. We observe all search sessions with at least one click on a fashion website,[7] as well as all other browsing activity within the session, including visits to non-fashion websites (e.g. checking email, visiting social networks, using search engines, etc). We define a "spell" as all sessions of a consumer before she makes a transaction, or before our observation period ends.[8] An observation in our data is a URL address of a website clicked by the consumer, together with a time stamp for the visit, and consumer demographics (e.g. age, gender).[9]

The data contain 7,877,551 total clicks and 427,768 fashion clicks. There are 4,612 consumers observed to search across 5,649 spells, purchasing a total of 3,017 fashion products, with 76% of spells containing no purchased product. As summarized in Table 1, on average, in a spell consumers visit 6 websites, make 75 clicks, look at 23 products, and spend 40 minutes searching. There are a total of nine fashion subcategories which were classified as (ordered by total purchases): "shirts, tops and blouses", "shoes", "pants and jeans", "underwear", dresses and skirts", "children's clothes", "jackets and vests", and "accessories". Zalando and H&M are the most popular websites visited across all subcategories.[10]

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

Insert Table 1 about here

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

A key feature of our data is that we observe clicks at a very granular level – we know the exact URL address of a website visited by the consumer for each click. This allows us to identify and differentiate clicks that come through the online advertising channel from other clicks that occur organically. We describe our method for detecting searches that are ad-initiated next.

[7]Consistent with the industry standard, GfK groups all clicks that are not interrupted by a time period of inactivity longer than 30 minutes into a "session."

[8]We note that most (62%) spells without a transaction end (have the last session) more than a week before the end of our observation period.

[9]Consistent with prior work on consumer search that utilizes a clickstream data set, we will treat a "click" as a "search" decision (e.g., De los Santos et al., 2012; Koulayev, 2014; Bronnenberg et al., 2016; Chen and Yao, 2017; Ma, 2016; De los Santos and Koulayev, 2017; Ursu, 2018; Gardete and Hunter, 2020; Ursu et al., 2020a).

[10]We refer the reader to (Ursu et al., 2020b) for further data descriptions. Appendix 10.1 contains details on additional data cleaning steps we performed.

3.2 Detecting Ad-Initiated Searches

We exploit the richness of the data contained in URL addresses to identify ad-initiated searches.[11] The main components of a URL are illustrated in the following example:

$$\underbrace{\text{https://}}_{\text{scheme}}\underbrace{\text{www.example.com}}_{\text{website}}\underbrace{\text{/path/index.html}}_{\text{path}}\underbrace{\text{?param=value}}_{\text{parameter}}\#\underbrace{\text{myfragment}}_{\text{fragment}}$$

A typical URL contains five components. First is the uniform resource identifier (URI) scheme, which for most websites is the http(s) communication protocol. Second, the URL identifies the website visited. Third, it identifies a hierarchical path representing different pages and subpages on the website. For example, the path will be empty if the consumer accesses the homepage of the website. Accessing a category page or a product page will then populate this path component. Fourth, the URL describes the parameters of the last path element identified. This could contain specifics of the page accessed or identify a query the consumer performed on the website. Finally, any other information is tracked using the fragment component.

We detect ad-initiated clicks using the parameter component of the URLs.[12] For clicks that are ad-initiated, the parameter component includes keywords that contain information about the advertiser (e.g. the brand that is advertising), the medium of advertising (e.g. email, display, social media), the ad agency (if any), and any other identifiers (e.g. specifics of the ad campaign). The most common URL parameters that identify advertisers contain a series of "UTM" (Urchin Tracking Module) keywords.[13] These parameters are standard tracking devices appended to a URL to allow marketers to track traffic across websites. There are five UTM parameters: "utm_source" identifying the advertising brand or the advertising agency employed (e.g. Copernica), "utm_medium" identifying the marketing medium (e.g. display, email), "utm_campaign" identifying an individual campaign name, slogan, promotion code, etc, "utm_term" identifying paid search keywords, and "utm_content" differentiating

[11] To classify the URLs, we first parsed the URLs using the R package called urltools (cran.r-project.org/web/packages/urltools/urltools.pdf). More information on how to parse URLs can be found at docs.python.org/3/library/urllib.parse.

[12] We note that there are at least two reasons for which our method may undercount the number of ads consumers search. First, consumers may be exposed to ads offline (e.g. on TV) and later search for those products online, a phenomenon that is well-documented (e.g., Joo et al., 2013, 2016). Such searches would be classified as active by our method. Second, a consumer may be exposed to an online ad, not click on it, and later return to search it on her own. Similarly, such searches would be classified as active.

[13] For more information on UTMs, see wikipedia.org/wiki/UTM_parameters or ga-dev-tools.appspot.com/campaign-url-builder. The latter reference shows how Google instructs advertisers to build an ad campaign using UTMs.

ad messages within the campaign. The "utm_medium" parameter also allows us to classify several types of advertising media, such as email, retargeting, search engine, display, etc.[14] The following are examples URLs with such UTM parameters highlighted in blue:

$$\underbrace{\text{https://}}_{\text{scheme}}\underbrace{\text{www.missetam.nl}}_{\text{website}}\underbrace{\text{/sale/page/906/}}_{\text{path}}$$

$$\underbrace{\text{?utm\_source=copernica\&utm\_medium=email\&utm\_campaign=20180315\_week11}}_{\text{parameter}}$$

$$\underbrace{\text{https://}}_{\text{scheme}}\underbrace{\text{www.missetam.nl}}_{\text{website}}\underbrace{\text{/sweater-print-blauw/product/98150/}}_{\text{path}}$$

$$\underbrace{\text{?utm\_source=criteo\&utm\_medium=retargeting\&utm\_term=126\&utm\_campaign=lowfunnel}}_{\text{parameter}}$$

In addition, we can identify clicks generated by online advertisements based on unique tracking parameters that search engines and platforms use. For example, "gclid", "gclsrc" and "dclid' are Google click identifiers, "msclkid" is a Microsoft Click Identifier, while "fbclic" is a Facebook click identifier. An example of such a URL is given below.

$$\underbrace{\text{https://}}_{\text{scheme}}\underbrace{\text{www.hunkemoller.nl}}_{\text{website}}\underbrace{\text{/nl\_nl/}}_{\text{path}}\underbrace{\text{?gclid=eaiaiqobchmisky9\&dclid=cpfiv9rfh9ocfuc}}_{\text{parameter}}$$

If a consumer navigates to a website organically the parameter component in the landing URL will not contain advertiser-related information. Instead, it can be empty, or contain information related to the consumer's search for product information (e.g. her query or her sorting and filtering options used). Below are a few example of such URLs:

$$\underbrace{\text{https://}}_{\text{scheme}}\underbrace{\text{www2.hm.com}}_{\text{website}}\underbrace{\text{/dames/jassen.html}}_{\text{path}}$$

$$\underbrace{\text{https://}}_{\text{scheme}}\underbrace{\text{www.plutosport.nl}}_{\text{website}}\underbrace{\text{/dames/zwemmen}}_{\text{path}}\underbrace{\text{?filter\_size=46\&manufacturer=speedo}}_{\text{parameter}}$$

[14]More information on how we classify such advertising types can be found in Appendix 10.2.

$$\underbrace{\text{https://}}_{\text{scheme}}\underbrace{\text{www.c-and-a.com}}_{\text{website}}\underbrace{\text{/nl/nl/shop/search}}_{\text{path}}\underbrace{\text{?q=sport}}_{\text{parameter}}\#\underbrace{\text{load-more-productsearch=24}}_{\text{fragment}}$$

In addition to providing us information on whether clicks are ad-initiated, the URL structure also allows us to identify website visits that are not related to acquiring (product) information directly. We defined clicks as "non-searches" if they included information on: transaction-related activities (e.g., add to basket, checkout, and order confirmation), login/out decisions, actions related to managing/viewing users' accounts or subscriptions, finding/creating a password, locating a store, tracking a shipment, or reaching customer service.[15] Below are two examples of non-search clicks.

$$\underbrace{\text{https://}}_{\text{scheme}}\underbrace{\text{www2.hm.com}}_{\text{website}}\underbrace{\text{/my-account/overview}}_{\text{path}}$$

$$\underbrace{\text{https://}}_{\text{scheme}}\underbrace{\text{www.missetam.nl}}_{\text{website}}\text{/}\underbrace{\text{orderdetails/page/880/}}_{\text{path}}\underbrace{\text{?orderid=3173951}}_{\text{parameter}}$$

These non-search clicks account for 10.20% of all consumer fashion clicks. The rest of the URL visits represent searches for product information, which will be the focus of our paper. We now describe the nature of these searches and the extent to which clicks are ad-initiated.

4 The Role of Online Advertising in the Consumer Search Process

4.1 Descriptive Analysis

Given the click classification presented above, we now describe the role that online advertising plays in the search process.

First, we find that ad-initiated searches correspond to a substantial fraction of overall website visits. Among all the search-related website visits, 15% are clicks on ads. Moreover, ads initiate the majority of website visits – 53% of first arrivals to a website are through clicks on ads.[16] This higher percentage

[15]Non-search clicks that are transaction related identify the products purchased. For more details, see (Ursu et al., 2020b).

[16]Here and throughout the paper, website visits are unique website-session combinations. If instead we considered unique website-spell combinations (thereby ignoring revisits), we would similarly find that 51% of website arrivals occurred through ads.

is due to the vast majority of within-website clicks occurring organically – not surprisingly, once on a website, consumers tend to navigate from page to page through links that are not sponsored.

The relative importance of ad-initiated searches varies across fashion subcategories and websites. Among subcategories, the "shoes" category has the largest fraction (approximately 16%) of clicks coming through the advertising channel, with "accessories" (13%) and "underwear" (approximately 10%) following closely behind. The "sweaters" category has the lowest share of ad-initiated clicks (just above 4%). These percentages are captured in Figure 1a. Among websites, the share of their total clicks that are ad-initiated varies widely, from 0 to more than 75%, with the average website having 12% ad-initiated clicks. Figure 1b displays these shares for the top 100 websites ranked by their total number of clicks. Although certain websites among the top, middle, and bottom ranked groups have a higher-than-average share of ad-initiated clicks, there is no clear pattern in these shares across website sizes.

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

Insert Figure 1 about here

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

Consumers are much more likely to click on ads early on in their search process. To illustrate this relationship, in Figure 2a we divide search spells into deciles and compute the average percentage of ad-initiated clicks in each decile.[17] We find that the share of ad-initiated clicks declines as search advances, such that for the first decile the share of ad-initiated searches is 22%, while for the last decile it is only 7%. This relationship is in large part due to the shorter within-website searches early on in the search process; the share of ad-initiated website visits is more stable throughout the spell, though still trending downwards towards the end of the search process, as shown in Figure 2b.

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

Insert Figure 2 about here

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

Using the parameter component of URL addresses, we can further classify the types of online ads that bring consumers to a website. We identify eight types of advertisements: affiliate (third party ads found on newspapers, blogs, etc), display, email, newsletter, retargeting, search engine, social, and

[17] To split searches into deciles, we follow the method used in (Bronnenberg et al., 2016). More precisely, if t denotes the search under consideration and N_i denotes the number of total searches performed in a spell i, then deciles are defined as $d(t,N_i) = ceil\left(\frac{10(t-r(0,1))}{N_i-1}\right)$, where $r(0,1)$ is a draw from a uniform distribution on the interval (0,1). Our results are robust to dividing searches into fewer or more than 10 groups, and to conditioning on spells with at least 3, 5, and 10 searches.

other.[18] In Figure 3, we show that the three most frequently clicked ad types are affiliate, email, and search engine, corresponding to 45%, 24%, and 16% of ad-initiated clicks, respectively. Classifying ad types also reveals that consumers were engaged in shopping-unrelated activities when they were exposed to ads – they were checking email, visiting social media websites, or reading the news. The importance of advertising types is relatively stable across stages of the search process – as captured by Figure A-1 in Appendix 10.5 – with email and search engine ad types becoming slightly less prevalent and affiliate ads becoming slightly more prevalent later in the search process.

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

Insert Figure 3 about here

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

Searches that are ad-initiated and those that are organic direct consumers to different landing pages. We separate out five categories of landing pages: homepage, listing page (e.g. a category page, such as "women's shoes", showing a list with several products on the same page), product page (a page dedicated to a single product), sales page, and other.[19] Most website visits coming through the ad channel land on homepages (32%) and listing pages (29%), with product pages, sales pages and other pages representing 9%, 3%, and 27% of clicks, respectively. The proportion landing page clicks coming through ads is substantially different from that occurring organically, where 45.3% of website visits start from the homepage, 26.6% start from a listing page, 12.5% start from the product page, 1% start from the sales page, and other pages account for the remaining 14.6%. These differences show that when search is ad-initiated, consumers are more likely to bypass the homepage and land directly on the listing or other pages. Furthermore, there are differences in landing pages across advertising types, as visualized in Figure 4. For instance, affiliate ads typical direct consumers to the homepage or the listing page, while email ads direct consumers to the listing and other pages most frequently. Retargeting ads almost exclusively bring consumers to product pages. Importantly, we note that none of the advertisement types generate the same proportions of landing page types as do clicks coming through the organic channel, suggesting that the nature of searches coming through the advertising channel is different. Clicks induced by search engine advertisements are most similar to searches coming from the organic channel, and might be counted as such, especially if consumers visit a search engine to search for a shopping-related keyword. We account for this possibility when estimating

[18] Appendix 10.2 provides more details on how we classified ads by medium.

[19] Examples of landing pages categorized as "other" include membership-exclusive pages or pages providing general fashion information, such as hm.com/nl_nl/life.

our model by running robustness checks and classifying clicks coming from search engines as active searches. More details can be found in Appendix 10.5.

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

Insert Figure 4 about here

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

Comparing the behavior of consumers on websites after arriving either through the ad or the organic channels, we find that the quality of ad-initiated website searches is generally lower. Table 2 summarizes differences between the two types of visits; ad-initiated website visits have on average fewer clicks (4.1 vs. 8.37 clicks), are shorter (2.28 vs. 4.32 minutes), and involve fewer products searched (0.58 vs. 1.28 products).[20] As a result, ad-initiated website visits are responsible for only 29% of transactions, while comprising 53% of website visits. Furthermore, in ad-initiated searches, consumers are exposed to more expensive products.[21]

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

Insert Table 2 about here

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

Consistent with the lower quality of ad-initiated website visits, consumers who heavily rely on searches through the advertising channel tend to make fewer clicks, search fewer products, and make fewer transactions. Table 3 describes search behavior for quantiles of consumers grouped by the share of searches they perform through the advertising channel. Consumers in the fifth quantile search predominantly through ads, with 85% of their searches being ad-initiated. These consumers make only an average of 24 clicks per spell, search for 13 minutes, inspect less than one product page, and only 1.7% of them have a transaction. In contrast, consumers in the second quantile, who click on a small number of ads (4%), have the longest search spells (153 clicks and 77 minutes) and inspect the most product pages (23). Also, 43% of them have a transaction. Interestingly, while search intensity monotonically decreases from the second to the fifth consumer quantile, in the first quantile search intensity is also lower – these consumers make only an average of 56 clicks and browse 10 product pages. We note that consumers in this group almost never click on ads, suggesting that they have a clear idea of the products they would like to search and purchase. Indeed, while these consumers search almost 3 times less than consumers in the second quantile, 31% of them have a transaction.

[20]All these differences are statistically significant at least at the 5% level, with Table 2 reporting the corresponding t-statistics.

[21]Since prices vary across fashion subcategories, we present in Table 2 the standardized price normalized by subtracting the average price and dividing by the standard deviation of the price in each subcategory.

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

Insert Table 3 about here

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

4.2 Relating Online Advertising with Passive Searches

Overall, the descriptive evidence presented above shows that product searches arriving through the advertising channel play an important role in the consumer search process – ad-initiated searches represent the majority of website visits, and thus need to be accounted for. In addition, these searches are systematically different from those occurring organically – they occur predominantly early in the search process, have a different composition of landing pages, a lower intensity of search after landing on a website, a higher propensity to click on more expensive products, and a higher propensity to leave without purchasing. These browsing patterns strongly suggest that searches coming through the advertising channel are different from the active search process that is assumed by standard search models. Below, we categorize them as passive searches, as opposed to organic clicks which we categorize as active searches.

Active searches are defined as effortful actions to seek out information about products relevant for a purchase decision (e.g., Weitzman, 1979). Such active search decisions involve determining which options to search, in what order to search them, and when to stop searching to make a purchase decision. Thus, we expect clicks coming through the organic channel to more accurately be described as active searches, since consumers decide which websites to search by typing in the information they are interested in – they navigate to a website by typing in its name or performing a search query.

In contrast, consumers do not have full control over the timing or the identity of the products they will be exposed to through ads. Instead, when exposed to ads, consumers merely choose how to react to them (click or not). To capture this difference, we treat ad-initiated clicks as **passive searches**, consistent with prior work (Renault, 2016). This categorization is further supported by the browsing patterns we presented above – while searches through online advertisements happen early on in consumers' search processes, they represent visits to lower quality websites, signaled by a lower intensity of searches on these websites, higher prices checked, and a lower likelihood to purchase. These patterns contradict the optimal search rules in Weitzman (1979), according to which consumers should sample the highest reservation utility options first, but fit the predictions of a model of passive search in which consumers merely react to information to which they are exposed. Furthermore, intuitively, consumers are unlikely to take the indirect route of checking email, visiting social media websites, or reading the news with the purpose of searching for information on fashion products, since searching fashion websites directly is easier. Thus, the fact that consumers click on ads when engaged in such shopping-unrelated activities is inconsistent with the notion of active search, which requires that consumers actively seek our product information. These highlighted patterns, as well as the additional differences between ad-initiated and organic searches discussed in the previous section, lead us to categorize ad-initiated clicks as passive searches, and organic clicks as active searches.

In the next section, we formalize the difference between ad-initiated and organic clicks by introducing a model of joint active and passive search.

5 Model

In the canonical sequential search model of Weitzman (1979), each search occasion the consumer decides whether to continue searching, in which case she chooses a product to search, or whether to stop searching, in which case she decides which product to purchase, if any. We refer to this type of search action as "active" since the consumer determines which product to search if search continues. In contrast, in a model of "passive" search, such as that of Renault (2016), consumers are assumed to search in response to firms' advertising, and thus to not be able to choose what product to search next. Instead, consumers observe an ad for a product and decide whether to obtain more information about it by searching. In what follows, we develop a model of sequential search where consumers make joint active and passive search decisions.

5.1 Setup: The Joint Active and Passive Search Model

Consider a consumer who is in the market for at most one unit of a product in a given product category. This consumer is aware of and is considering towards her next purchase options $j \in J$. The consumer is uncertain about the options available to her, but may resolve that uncertainty by searching. Searching is costly, $c_j > 0$, but reveals a potential payoff u_j drawn from a distribution function $F_j(\cdot)$ with support $[\underline{\theta}_j, \bar{\theta}_j]$. At each decision point, the consumer has searched a set of products S, and a set $\bar{S}$ is available to search, where $S \cup \bar{S} = J$. Let the maximum reward observed among the searched options be given by

$y = \max_{j \in S \cup \{0\}} u_j$, where $j = 0$ denotes the outside option of not purchasing. At the end of the search process, the consumer may choose to purchase one of the options searched, or may choose the outside option. This consumer solves the following problem (due to Weitzman (1979))[22]

$$V(\bar{S}, y) = \max_{stop,continue} \{y, \max_{j \in \bar{S}} -c_j + W_j(\bar{S}, y)\}, \tag{1}$$

where $V(\emptyset, y) = y$ and the continuation value $W_j(\cdot)$ for $j \in \bar{S}$ is given by

$$W_j(\bar{S}, y) = V(\bar{S} \setminus j, y) F_j(y) + \int_y^{\bar{\theta}_j} V(\bar{S} \setminus j, u) dF_j(u). \tag{2}$$

In words, at a given moment in the search process, the state space describing the problem of the consumer is given by the set of options she is aware of and is considering, but has not yet searched, $\bar{S}$, and by the best option revealed so far, y. At that moment, the consumer may decide to stop searching and choose the best option revealed so far, y. Alternatively, the consumer may choose to continue searching, in which case she searches one of the options in $\bar{S}$.

In addition to the J options a consumer is aware of, advertisers may inform the consumer about options $a \in A$. This modeling choice is consistent with prior work showing that the primary mechanism through which advertising affects the consumer search process is awareness (Goeree, 2008; Terui et al., 2011; Honka et al., 2017; Tsai and Honka, 2018), and with the literature documenting the informative effects of advertising more broadly (Ackerberg, 2001, 2003; Abhishek et al., 2012; Blake et al., 2015; Sahni and Zhang, 2020). When a consumer is exposed to an ad from a, her choice is

$$V(\bar{S} \cup a, y) = \max_{stop,continue} \{y, \max_{k \in \bar{S} \cup a} -c_k + W_k(\bar{S} \cup a, y)\}, \tag{3}$$

mirroring equation 1 with a now in the awareness set. We follow Renault (2016) and assume the consumer is passive in her reaction to ads. That is, the consumer does not have control over the probability of observing an ad, the identity of the advertiser, or the timing of the ad (in contrast to active search where she is assumed to choose an optimal search order for all options available). Thus, the consumer does not change her search process in anticipation of the arrival of ads. However, when exposed to the ad, the consumer decides how to react to it. This is one variation of our model of active

[22]We assume no time discounting, consistent with prior empirical work, (Kim et al., 2010, 2017; Chen and Yao, 2017; Honka and Chintagunta, 2017; Ursu, 2018; Ursu et al., 2020a).

and passive search, to which we will refer as the **AP-weak model**.

While the AP-weak model captures the lack of control of consumers over the arrival of advertised product options, it still assumes that consumers compare a to the rest of the products not yet searched, $\bar{S}$. However, the consumer might be engaged in other online activities when exposed to ads (e.g. checking email, visiting social networking websites, etc) and may not consider other options at that moment (Renault, 2016); or the ad might focus the consumer's attention and limit her ability to process information about other options due to cognitive constraints (Gossner et al., 2020). To account for these possibilities, we propose a stronger version of the active and passive search model, in which the consumer does not compare the advertised option to other unsearched options when exposed to the ad, and instead compares the ad only with the best option searched so far, y, as per

$$V(a,y) = \max_{not\ a,\ a} \{y, -c_a + W_a(a,y)\}. \tag{4}$$

Since ads are not anticipated, we assume the consumer can always continue the search process (solve equation 1) after deciding whether to search ad a or not. We refer to this model as the **AP-strong model**.[23]

We highlight several important assumptions behind the AP-weak and AP-strong models. First, we follow the results in Terui et al. (2011) and Honka et al. (2017) and assume that advertising affects the consumer's choice process at the awareness stage. Alternatively, ads can affect the consumer's match value (e.g., Hastings et al., 2017), which would shift the realized y for the advertised products in our model. Since we focus on consumers' search for information, we assume that ads affect consumers' navigation to webpages, and only through that affect their expected utility, price or feature knowledge.

Second, we assume that consumers do not anticipate ad exposure and cannot affect the probability or the timing of observing a specific ad, since this would contradict the notion of passive search. If instead consumers could affect the probability or the timing of ads, then it might be more reasonable to treat these searches as active searches as well. To test our assumption, in the empirical part of the paper we also estimate the Weitzman (1979) model on our data – a model that assumes consumers are aware of all products (i.e. $A = \emptyset$) and that all search is active. However, we find that both of our models of advertising as passive search outperform the Weitzman model of purely active search.

[23] Note that this model is designed to capture consumers' suboptimal decisions in response to ads, for the reasons explained above. Thus, the value function $V(a,y)$ solely captures the main tradeoff consumers make, not their entire continuation value for each decision.

Finally, we do not model firms' advertising strategic decisions, since they are not observed in our data.[24] In reality, an important aspect of online advertising is that firms can target ads to a specific consumer, based on, for example, her previous search history. Part of this targeting might result in ads being shown to consumers when they are likely to click, similar to scenarios considered in Blake et al. (2015) and Simonov et al. (2018). Such precise advertising targeting should make ad-initiated searches look more like active searches, something we do not find support for in our data – consumers' ad-initiated searches are typically short, involve fewer and more expensive products, and are unlikely to lead to a purchase. Nevertheless, we account for this possibility by estimating the Weitzman (1979) model that treats all searches as active.

5.2 Search Rules

Having laid out the primitives of the joint active and passive search model, we now describe the optimal search rules.

In the absence of ads, the AP-weak and the AP-strong models coincide with the Weitzman model. For this problem, the optimal search strategy is given by the following search rules:

1. **Selection rule:** If a search is to be made, then the option $j^* \in \bar{S}$ with the highest reservation utility z_{j^*} should be searched next, where

$$c_{j^*} = \int_{z_{j^*}}^{\bar{\theta}_{j^*}} (u - z_{j^*}) dF_{j^*}(u). \tag{5}$$

2. **Stopping rule:** Search should terminate when the maximum utility observed so far exceeds the reservation utility z_{j^*} of any unsearched option.

3. **Choice rule:** Once search has terminated, the option with the highest revealed utility among those searched (including the outside option) should be chosen.

In words, if the consumer is not exposed to an ad, then she will search using Weitzman's search rules. That is, among the options available to search $\bar{S} \subset J$, the consumer will rank products by their reservation utilities and continue searching if there exists an option j^* with reservation utility larger

[24] We note that we share this data limitation and modeling approach with prior empirical work on consumer search that uses clickstream data (Kim et al., 2010, 2017; Koulayev, 2014; Chen and Yao, 2017; Ma, 2016; Honka and Chintagunta, 2017; De los Santos and Koulayev, 2017; Ursu, 2018; Gardete and Hunter, 2020; Ursu et al., 2020a). We add to the literature the idea that some of the clicks consumers make do not result from an active search process, so should be dealt with differently.

than the highest utility observed so far, i.e. if $z_{j^*} \geq y$ and $z_{j^*} \geq z_j, \forall j \in \bar{S}$.[25] When the highest utility observed through search exceeds the reservation utility of any option not yet searched, the consumer stops searching and makes a purchase decision.

If instead the consumer is exposed to an ad a in the AP-weak model, then she solves a problem very similar to the one in Weitzman, except that another option has been added (exogenously) to the set of available options to search. Thus, the consumer will search ad a if $z_a \geq z_j$ and $z_a \geq y, , \forall j \in \bar{S}$. In contrast, in the AP-strong model the consumer will search ad a if $z_a \geq y$.

We will estimate four different models on our data.

1. The **AP-weak model**: The joint active and passive search model proposed above where $A \neq \emptyset$, and the consumer compares z_a with the utility of all options searched so far, y, *and* with the reservation utility of all options not yet searched in $\bar{S}$.

2. The **AP-strong model**: A stricter version of the joint active and passive search model proposed above where $A \neq \emptyset$, and the consumer compares z_a only with the utility of all options searched so far, y.

3. The **Weitzman model**: the case where $A = \emptyset$.

4. The **Weitzman model with advertising costs**: the case where $A = \emptyset$, and ads affect the search cost of the consumer.

5.3 Example Illustrating Consumer Search Rules Across Models

To better illustrate differences between the search rules in the four models above, consider the example provided in Table 4. Suppose there are five products and an outside option available, that the consumer has searched three of them, and that options 2 and 4 exposed the consumer to ads. Also, suppose that only the ad for option 2 was searched. We ignore the choice rule in this example since it is the same across all models.

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

Insert Table 4 about here

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

[25] Without loss of generality, we assume throughout the paper that when the consumer is indifferent between searching and stopping search, she will continue searching, and that when the consumer is indifferent between buying and choosing the outside option, she will choose to buy.

Columns (i), (ii), and (iii) in Table 4 describe the restrictions on the parameters of interest imposed by the Weitzman, AP-strong, and AP-weak models, respectively. Compared to the Weitzman model, the AP-strong model assumes the consumer is not aware of options 2 and 4. Therefore, their reservation utilities do not affect the search order. Rather, when the consumer is exposed to an ad for these two options, she decides whether to search them solely by comparing their reservation utility with the utility of options searched so far. The AP-weak model maintains this assumption, but also models the consumer as comparing the reservation utility of the advertised options to the reservation utility of unsearched options. The fourth variation of our model, the Weitzman model with advertising costs, is not illustrated in this example, but would impose the same restrictions on the reservation and revealed utilities as the Weitzman model. In this variation, ads affect consumers' search costs, rather than affecting their awareness, as in the AP models.

5.4 Related Problems

Our proposed model of joint active and passive search is related to three problems found in the literature. First, the literature on "arm-acquiring bandits", pioneered by Whittle (1981), considers an extension of the traditional multi-armed bandit problem where arms appear continually while the decision maker evaluates them. Whittle (1981) shows that by assigning a state to every arm, the Gittins index solution that applies in the multi-armed bandit problem continues to hold for i.i.d. arrivals of the arms. The decision rule dictates that the decision maker operate the arm in the state with the largest index as long as it is higher than the best observed reward so far; otherwise the decision maker should stop the process and exploit the best arm. Our model is related to this problem if we think about ads as such arms that are added to the problem the consumer is solving. The difference is that in our case, ads do not appear continually and the consumer does not take their arrival into account.

A second problem related to our model is that of endogenous awareness sets, studied in two recent papers (Greminger, 2020; Fershtman and Pavan, 2020). In these papers, the consumer has the choice to search among options she is aware of or to discover new options that are then added to her awareness set. The authors then describe conditions under which an index policy exists. Our model is similar in the sense that we assume consumers are not aware of or are not considering the options to which ads expose them. However, in our model consumers do not choose to expand their awareness set; rather, ads arrive exogenously, expanding this set.

Finally, our model is also related to the rich literature on random search (e.g., Wolinsky, 1986), where consumers do not choose the order in which they search, but merely choose when to stop searching. Our model relaxes the assumption that consumers choose the order of search for the ads they observe, in the spirit of random search models.

6 Empirical Application and Estimation

6.1 Empirical Model

In our empirical application, we model consumers as searching across websites (e.g., adidas.com, nike.com) in one of the four largest fashion subcategories in our data: (1) shirts, tops, and blouses, (2) shoes, (3) pants and jeans, and (4) underwear.[26] Appendix 10.3 provides details on how the estimation samples were constructed. In the model, consumer $i = 1,\dots,N$ seeks to purchase from website $j = 1,\dots,J$ or to choose the outside option of not purchasing, $j = 0$. Consumer i's utility of purchasing from website j is given by

$$\begin{aligned} u_{ij} &= v_{ij} + \epsilon_{ij} \\ &= w_j + \gamma X_{ij} + \eta_{ij} + \epsilon_{ij} \end{aligned} \quad (6)$$

where v_{ij} denotes the information the consumer has about a website before searching it, and ϵ_{ij} denotes the information revealed through search. The information on v_{ij} includes website intercepts, w_j, observed website and consumer characteristics, X_{ij} – such as measures of website loyalty (the number of times the consumer has previously searched a website in other subcategories or in previous spells) and price sensitivity (whether the consumer visited the sales page of a website) – and characteristics unobserved by the researcher but observed by the consumer before search, η_{ij}. Since in the fashion industry prices do not vary over a short time period or across consumers, they do not affect consumers' utility after controlling for website fixed effects.[27] We assume that both η_{ij} and ϵ_{ij} are distributed as standard normal distributions (consistent with prior work (Kim et al., 2010; De los Santos et al., 2012; Honka, 2014; Chen and Yao, 2017; Honka and Chintagunta, 2017; Ursu, 2018; Ursu et al., 2020a)).

[26]We choose to model search across websites (rather than products within websites) for several reasons: (i) consumers are more likely to search websites directly rather than their individual product subpages, since they rarely know which products are available before navigating from the homepage to various list pages that display such products; (ii) ads vary in the types of landing pages to which they direct consumers, so this modeling assumption allows us to keep our analysis consistent across ads; (iii) developing a model of search across as well as within websites is beyond the scope of our paper.

[27]Prices vary mostly across seasons (e.g. when companies run sales promotions) and are generally not personalized to individual consumers. For more facts about pricing in the fashion industry, we refer the reader to Ursu et al. (2020b).

The outside option does not require searching and has a utility equal to $u_{i0} = q_0 + \eta_{i0}$, where q_0 is an intercept denoting the value of not purchasing.

Searching to resolve uncertainty about ϵ_{ij} is costly for consumers. Search costs are given by $c_{ij} = \exp(\kappa)$, modeled as exponential functions to ensure that they are positive and consistent with prior work (e.g., Honka, 2014; Chen and Yao, 2017; Ursu, 2018). In the Weitzman model with advertising costs we allow for the possibility that ads have different search costs, $c_{ia} = \exp(\kappa + \delta Ad_{ia})$.

In our data, we only observe ads that consumers have clicked. We use our data to impute the probability of ad exposure based on the search history of the consumer, as described in Appendix 10.3. This approach allows us to more accurately capture the magnitude of the effect of passive search and to be able to consider the Weitzman model with advertising affecting search costs. However, this assumption does not drive our results – our estimates are robust to using only ad clicks and not ad exposures (see Table A-2 in Appendix 10.5).

6.2 Estimation

The four model variations we will estimate with our data are based on the Weitzman (1979) model. Therefore, we will first describe the estimation procedure of the Weitzman (1979) model that is commonly used in the literature (Kim et al., 2010; Chen and Yao, 2017; Honka and Chintagunta, 2017; Ursu, 2018), and then describe how the other variations differ.

In the Weitzman (1979) model, consumers search options 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 search rules described in Section 5.2 above translate into the following restrictions on preferences and search cost parameters.

Suppose consumer i searched a number s of websites and that she chose j after stopping her search (including the outside option). With a slight abuse of notation, we order websites by their reservation utilities and let n denote the website with the nth largest reservation utility. Since consumers searched websites by reservation utilities, according to the *selection rule*, it must be that[28]

$$z_{in} \geq \max_{k=n+1}^{J} z_{ik}, \qquad \forall n \in \{1, \ldots, J-1\}. \tag{7}$$

In addition, the *stopping rule* imposes the following two restrictions. For the set of websites searched,

[28]For details on how to compute reservation utilities in our setting, we refer the reader to Kim et al. (2010).

it must be that

$$z_{in} \geq \max_{k=0}^{n-1} u_{ik}, \qquad \forall n \in \{1, \ldots, s\}. \tag{8}$$

In contrast, for the websites that were not searched, it must be that

$$z_{im} \leq \max_{k=0}^{s} u_{ik}, \qquad \forall m \in \{s+1, \ldots, J\}. \tag{9}$$

Finally, consistent with the *choice rule*, if the consumer chooses j (including the outside option), then her utility from this choice is larger than that of any other searched website, i.e.,

$$u_{ij} \geq \max_{k=0}^{s} \ u_{ik}, \qquad \forall j \in \{0, 1, \ldots, s\}. \tag{10}$$

In what follows, we describe how each model we estimate varies from the Weitzman setup.[29]

1. The **AP-weak model**: for all websites, this model uses the same equations (7-10), except that the set of options in equation (7) does not include any of the advertised websites unknown when searching n.

2. The **AP-strong model**: (i) for websites that did advertise, this model does not impose equation 7; and (ii) for all other websites, this model uses equations (7-10), except that the set of options in equation (7) does not include any of the advertised websites unknown when searching n.

3. The **Weitzman model**: no variation.

4. The **Weitzman model with advertising costs**: this model uses the same equations (7-10) with search costs as a function of advertising.

Differences in the selection and stopping rules across these models are illustrated in the example in Section 5.3.

In addition to the restrictions imposed in equations (7-10) and their variations, we assume that the first search performed by a consumer is free.[30] This assumption is common in prior work (e.g., Honka, 2014; Honka and Chintagunta, 2017) and is necessary since all consumers in our data search at least once.

If consumers search using the rules described above (equations (7-10) and their variations), then they make search and purchase decisions jointly. Thus, the probability of observing a certain outcome

[29] For identification purposes, we do not consider versions of the AP-weak or AP-strong models with advertising affecting search costs.

[30] We allow for the possibility of no search in our Monte Carlo simulation in Section 6.5.

in the data for consumer i is characterized by the joint probability of equations (7-10) holding. This probability is given by

$$L_i = Pr\Big(\text{Selection rule}_i,\ \text{Stopping rule}_i,\ \text{Choice rule}_i\Big). \tag{11}$$

Because consumers make these decisions jointly, the likelihood function does not have a closed-form solution. We use a simulated maximum likelihood approach to estimate the parameters of the model. In choosing the simulation method, we use the logit-smoothed AR simulator following the previous literature McFadden (1989); Honka (2014); Honka and Chintagunta (2017); Ursu (2018); Ursu et al. (2020b). Implementation details are discussed in Appendix 10.4.

6.3 Identification

Parameter identification in the four models we estimate follows from the identification argument used by standard consumer search models based on the Weitzman model (Kim et al., 2010; Chen and Yao, 2017; Honka and Chintagunta, 2017; Ursu, 2018). More precisely, 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. In addition, variation in the frequencies with which websites are searched first, second, etc will further pin down website intercepts. These same data patterns together with the selection, stopping, and choice rules described in Section 6.2 help recover preference estimates of advertising websites in all the models we consider.

Similarly, as in prior work, search costs do not affect purchase decisions (i.e. do not enter the choice rule) and are identified from the number of websites that consumers search. More precisely, the search rules impose an upper and a lower bound on the search cost parameter κ 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 expression of the reservation utility. Finally, for the Weitzman model with advertising costs, the parameter δ shifting search costs due to ads is identified from variation in which and how many ads consumers search and buy.

6.4 Biases in Parameter Estimates

The model variations we consider describe consumer search decisions differently, leading to different parameter estimates. In this section, we describe how these parameter estimates compare with those from the Weitzman model. As we will show below, the two active and passive search models we consider, as well as the Weitzman model with advertising affecting search costs, outperform the standard Weitzman model. Therefore, we will interpret these differences in parameter estimates as biases due to misspecifications of the Weitzman model.

Consider first the AP-weak model. This model does not require the reservation utilities of ads to be lower than the reservation utilities of options searched before them. Therefore, it allows for the possibility that ads have higher reservation utilities (higher expected utilities and lower search costs) than those in the Weitzman model. More precisely, the AP-weak model allows for the possibility that advertising websites were not searched because consumers were unaware of them, not because consumers actively chose not to search them. If this model coincided with the true data generation process, then compared to it, the Weitzman model would underestimate the expected utility of frequently advertised websites. Given these underestimated expected utilities, search cost estimates in Weitzman may remain unchanged or may be slightly higher than in the true AP-weak model. Finally, the outside option estimate would be underestimated by the Weitzman model to rationalize consumers' decisions not to purchase when advertised websites have lower expected utility estimates.

Next, consider the AP-strong model. In addition to not requiring reservation utilities of ads to be lower than the reservation utilities of the options searched before them (as in the AP-weak model) this model also does not require that reservation utilities of ads be higher than the reservation utilities of options not yet searched. With both a lower bound and an upper bound on reservation utilities removed, this model may lead to either higher or lower estimates of consumers' valuation for advertised websites compared to the Weitzman model. The direction of the bias depends on the timing of ad-initiated searches. If the advertised websites are searched predominantly early in the search process – as we generally observe in our data – the lower bound on reservation utilities would be relatively more important in affecting estimates, since many options are yet to be searched. This should lead to an upward bias in the expected utility estimates of advertised websites in the Weitzman model, which incorrectly imposes that websites searched early have higher reservation utilities than those searched later or those not searched. In contrast, the Weitzman model will underestimate the expected utilities of advertised websites if they are searched predominantly later in the search process. Biases in search costs and the outside option parameters resemble those in the AP-weak model.

Finally, consider the Weitzman model with advertising affecting search costs. Consistent with the data patterns we presented in Section 4, advertised websites are rarely purchased, but frequently searched. Thus, the Weitzman model with constant search costs will likely overestimate advertised websites' utilities, but estimate a negative effect of advertising on search costs.

6.5 Monte Carlo Simulation

We now show that the simulated maximum likelihood method using the logit-smoothed AR simulator can recover the parameters of our model. We do so using Monte Carlo simulations. We generate a data set of 1,000 consumers making choices among five options – four websites and an outside option. We simplified the model estimated to include only website intercepts, an outside option intercept, and a mean search cost parameter. The true values of these parameters are similar to those from a preliminary estimation of our model. Website 4 will serve as the reference option.

To determine how the presence of ads affects estimates, we choose website 2 as the advertiser. With 25% probability, consumers are aware of website 2 and search it according to the Weitzman optimal search rules. All other consumers are not aware of this website but are exposed to its ad. This assumption allows us to better mimic our data where the same website may be searched organically by some consumers, and through an ad by others. However, the results we present below would continue to hold under different scenarios, including in the case where website 2 advertises to all consumers.

We varied the timing of ad exposure as follows: consumers searched websites other than website 2 in decreasing order of their reservation utilities; website 2 had a temporary value for its reservation utility equal to the average reservation utility in the data. Therefore, the ad sometimes appeared before the consumer searched any other options, other times it appeared after the last searched option, but most often it appeared somewhere in the middle. When exposed to the ad, consumers chose whether to search it or not.

To estimate our model, we follow the steps described in Section 6.2 and Appendix 10.4 and use 500 draws from the distribution of utility error terms for each consumer-website combination to construct the likelihood function. We repeat the estimation on 50 different data sets generated using the same true parameters, but different seeds for the utility errors terms.

Our Monte Carlo simulation results are displayed in Table 5. In column (i), we present the true parameters; in column (ii), we show results when data were generated according to the AP-strong model; in column (iii) we show results when data were generated according to the AP-weak model. For each set of data we generated, we estimate two models: the corresponding AP model and the Weitzman model. The coefficients reported represent averages across 50 estimations of our model. In parentheses, we also report the standard deviation of these estimated coefficients.

Two findings are worth emphasizing. First, we find that each version of the AP model, when used to estimate parameters on data that it generated, can recover those parameters well. Second, the Weitzman model, when used to estimate parameters on data generated by either version of the AP model, recovers a biased estimate of the advertiser's value, with less or no bias for other parameters. More precisely, the Weitzman model underestimates the value of the advertiser, confirming the predictions from Section 6.4. In this simulation, we focused on a simple model with one advertiser. These effects would be inflated when most or all websites advertised. Also, by exposing consumers to the advertising website predominantly early on in the search process (rather than randomly in the current setup), the Weitzman model would overestimate the value of the advertiser in the AP-strong model, but not in the AP-weak model.

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

Insert Table 5 about here

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

7 Results

7.1 Estimation Results

We estimate our models on four different fashion subcategories, "shirts, tops, and blouses", "shoes", "pants and jeans", and "underwear". Table 6 below presents our results from the first two subcategories, while Table 7 presents results from the other two subcategories. In bold, we identify the three largest advertisers in each subcategory.

To start, we describe the overall takeaways from our estimation results, consistent across all models and subcategories. As expected, we find that Zalando and H&M, the two largest fashion retailers in the Netherlands, are among the top favorite websites for consumers across several categories. All else equal, consumers prefer websites they visited before – in other subcategories or in previous spells – and visiting a price discount page corresponds to a higher indirect utility, potentially signaling consumers' price sensitivity. The search cost estimates are positive and the coefficients are significant, indicating that consumers get disutility from search. The magnitude of the search costs estimates implies that a 10% increase in search costs per website would decrease total searches by approximately 2%.

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

Insert Table 6 about here

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

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

Insert Table 7 about here

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

We now turn to comparing the model estimates. In all subcategories, the main differences in the estimates come from the advertised websites' utilities, highlighted in bold. The estimates of utilities of other websites are statistically similar across the models.

First, consider the estimates of advertised websites' utilities in the AP-strong (column (i)) and the Weitzman (column (iii)) models. In subcategories 1, 3, and 4, the AP-strong model has on average 18% lower estimates of advertised websites' utilities compared to the Weitzman model. Since in all of these subcategories the vast majority of advertised websites are visited early on in the search process (as shown in Table A-1 in Appendix 10.3), our result is consistent with the expected upward bias of the Weitzman model described in Section 6.4. Search cost estimates are 15-30% higher in subcategories 1, 3 and 4 in the AP-strong model than in the Weitzman model. In subcategory 2, the bias of the Weitzman model goes in the opposite direction, with the advertised websites' utilities having on average a 10% higher estimate in the AP-strong model. This result similarly aligns with our expectations, since advertising websites in subcategory 2 are more often searched later in the search process (see Table A-1 in Appendix 10.3).

Second, consider the AP-weak model (column (ii)). The estimates of advertised websites' utilities in this model are higher than in the Weitzman model across all subcategories, as predicted in Section 6.4. For example, in subcategory 1, About You, C&A, and Debijenkorf have 4-10% higher estimated website intercepts in the AP-weak model than in the Weitzman model.

Finally, in the Weitzman model with advertising costs (column (iv)), we find that the estimates of advertised websites' utilities are on average 25% lower compared to the Weitzman model, but it is less costly for consumer to search these options, since δ estimates are negative.

Three measures of model fit suggest that the AP-strong model is the most appropriate for our data, followed by the AP-weak model and the Weitzman model with advertising costs. The bottom panel of each table lists the measures of model fit we computed. First, the AP-strong model has the lowest log-likelihood, followed by the AP-weak model, and lastly by the two variations of the Weitzman model. However, we cannot directly compare the likelihood values since the likelihood functions across models are different. Instead, we rely on two other measures of fit, the mean absolute distance (MAD) and the root mean squared error (RMSE) in the purchase and search shares of each website. We compute these measures by taking a model's parameter estimates and simulating consumer choices within sample, averaging out the effect of the utility error terms across 50 simulations. We finally compared the predicted and the observed purchase and search shares in our data to compute the MAD and the RMSE.

Across all subcategories, we find that the AP-strong model predicts purchases better than all other models (has lower MAD and RMSE values), followed by the AP-weak model, and the Weitzman model with advertising costs. This is particularly remarkable since the AP models use fewer inequalities in estimation than the Weitzman model, leaving more degrees of freedom unused. We note that the AP models cannot predict searches of advertising websites since consumers are assumed to be exposed to ads exogenously. Therefore, to compare model fit on search shares, we report the predicted search shares only for websites searched actively under each model. We again find that the AP-strong model outperforms all other models, generally followed by the AP-weak model and the Weitzman model with advertising costs.

We also highlight that the AP-strong model can explain the data patterns we presented in Section 4.2 better than all other models considered. For example, the AP-strong model can explain why ads are more likely to be clicked early rather than late in the search process (Figure 2). This is the case because early in the search process, the best option searched so far has a relatively low value compared to its value later in the process. Therefore, even low quality ads are more likely to be clicked. In contrast, neither the AP-weak nor the Weitzman models can rationalize this pattern, since they impose the constraint that advertised websites need to have higher reservation utilities than all options searched after them. Similarly, the lower advertised websites' utilities recovered by the AP-strong model are consistent with the lower quality of the ad-initiated searches described in Section 4.2 – these searches involve fewer and more expensive products, are shorter, and are less likely to lead to a purchase. All these facts support the idea that the AP-strong model is a useful model through which to understand the role of online advertising as passive search. Our results also provide evidence of consumers' limited ability to compare advertised options to other unsearched products, an important behavioral limitation that models of active search ignore.

7.2 Simulated Substitution Patterns

To understand how advertising affects competition between websites, we use our model's estimates to examine consumer substitution patterns under different advertising scenarios. Towards this end, we use the AP-strong model estimates (Tables 6 and 7) to simulate consumer search and purchase decisions under different advertising scenarios, and then compare the resulting outcomes to the current benchmark. We repeat the simulation 50 times and report the mean results to integrate over the distribution of unobserved utility shocks. Consistent with our model, we assume that consumers are not aware of advertising websites if ads are shut down, absent the crowd-out behavior described in Blake et al. (2015) and Simonov et al. (2018). Table 8 presents overall changes in the volume of searches and purchases across subcategories, and Figure 5 presents the breakdown of these changes by websites for the first two subcategories.[31]

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

Insert Table 8 about here

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

We first examine a scenario in which all advertising is shut down. In the first panel of Table 8, we present the change in total searches and purchases in each subcategory. When there is no advertising, the model predicts that consumers perform approximately 20-36% fewer searches and, as a result, make 11-21% fewer purchases. This is driven by large decreases in searches and purchases by websites that advertise frequently, as shown in the first panel in Figure 5 – the top three advertisers experience about a 60% decrease in their searches and purchases. This change highlights the importance and the prominence of paid traffic for the largest advertisers in our data. Other websites experience either a small decrease or an increase in their traffic, due to consumer substitutions in their search and purchase decisions.

[31] Our findings in Figure 5 generalize to the remaining subcategories and are available from the authors upon request.

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

Insert Figure 5 about here

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

The second scenario considers the case where only the three largest advertisers cannot advertise, while all other websites continue to expose consumers to ads at the same time as in our data. Once again, we find lower total searches and purchases, albeit the change is smaller: there are 11-17% fewer searches and 2-6% fewer purchases. The second panel of Figure 5 shows a much more uniform substitution of consumers to websites that continue to advertise – these websites experience 10-40% more total searches and purchases.

Finally, in the third scenario we consider the case where advertising by the three websites with the highest number of purchases in each subcategory is shut down. In this case, the decrease in searches and purchases is relatively small – 2-11% for searches and 2-8% for purchases – since these websites rely less on advertising to generate searches and transactions. As expected, websites that stop advertising experience a decrease in searches and purchases, as consumers substitute towards other websites, which experience a change in the opposite direction.

These results show how our model can be used to examine the role of advertising in generating searches and purchases for websites. We note that these results rely on the awareness mechanism behind the effect of advertising, demonstrated in prior work (Goeree, 2008; Terui et al., 2011; Honka et al., 2017; Tsai and Honka, 2018). This mechanism fits naturally within the information search process of a consumer, which is the focus of our study. Consumers' lack of awareness of advertised websites leads to a decrease in searches and, ultimately, to a decrease in purchases.

8 Managerial Implications

Our results describe the role of ad-initiated clicks in the consumer search process – ad-initiated searches represent the majority of website visits, occur predominantly early in the search process, and are unlikely to lead to a transaction. Not taking this into account will distort estimates of consumer preferences and, ultimately, managerial decision making. For example, assuming a consumer has actively searched a product on Nike.com, rather than passively reacted to a Nike ad – even if it is for the same product – implies wrongly assuming that a consumer expects Nike's product offerings to dominate those of other brands, inflating consumer brand preferences.

For advertisers, understanding the extent to which consumers seek out their products actively rather than only react to their product messages, can help inform advertising decisions. Our model is tailored to measure the degree to which consumers will substitute away from searching the advertised products towards searching competitors' products, further improving advertisers' strategies. As a result, consumers may benefit from companies' better understanding of their preferences through more relevant ads or better fitting products.

More broadly, our paper questions the assumption that every click performed by a consumer online is an outcome of an active search process. Beyond the case we focus on in this paper, where advertising affects search decisions, there are several other cases in which we expect a click to not reflect an active search decision. For example, in many settings, consumers may merely be curious about a product or may be browsing rather than searching for information with the goal of making a purchase decision (e.g. a consumer typing in "Ferrari" into Google out of curiosity, rather than because she is interested in gathering information towards her next purchase). In such cases, companies should account for passive searches when running their (re-)targeting advertising campaigns – if a consumer stumbled upon a product webpage while browsing, it might be a weak signal of the consumer's interest in buying the product and it may thus be wasteful to (re-)target this consumer with online advertising. A broader understanding of passive search settings (perhaps using similar methods as Moe (2003)), as well as a formal treatment of decision making in such cases would be theoretically and managerially relevant.

9 Conclusion

In this paper, we examine the role of online advertising in consumers' information search process. Using a detailed clickstream data set capturing website visits at the exact URL level, we develop and apply a method that classifies clicks into ad-initiated and organic searches. We then show that ad-initiated searches are extensive – driving more than half of all website arrivals – happen early on in the search process, and tend to lead to less in-depth and overall lower quality searches. These patterns do not align with standard models of active information search (Weitzman, 1979), and instead are consistent with models of passive search, such as Renault (2016). To account for such passive search, we develop a simple model that accommodates both active and passive search decisions by consumers and estimate this model on the four largest fashion subcategories in our data. The results show that a model of active and passive search fits the data the best, while treating all searches as active leads to substantial biases in the estimates. We use the model estimates to describe the substitution patterns of consumers' search and purchase decisions in the absence of online advertising.

References

Abhishek, V., Fader, P., and Hosanagar, K. (2012). Media exposure through the funnel: A model of multi-stage attribution. Working paper.

Ackerberg, D. A. (2001). Empirically distinguishing informative and prestige effects of advertising. *RAND Journal of Economics*, pages 316–333.

Ackerberg, D. A. (2003). Advertising, learning, and consumer choice in experience good markets: An empirical examination. *International Economic Review*, 44(3):1007–1040.

Anderson, S. P. and Renault, R. (1999). Pricing, product diversity, and search costs: A Bertrand-Chamberlin-Diamond model. *The RAND Journal of Economics*, 30(4):719–735.

Anderson, S. P. and Renault, R. (2006). Advertising content. *American Economic Review*, 96(1):93–113.

Anderson, S. P. and Renault, R. (2013). The advertising mix for a search good. *Management Science*, 59(1):69–83.

Beales, H., Mazis, M. B., Salop, S. C., and Staelin, R. (1981). Consumer search and public policy. *Journal of Consumer Research*, 8(1):11–22.

Beatty, S. E. and Smith, S. M. (1987). External search effort: An investigation across several product categories. *Journal of Consumer Research*, 14(1):83–95.

Bennett, P. D. and Mandell, R. M. (1969). Prepurchase information seeking behavior of new car purchasers—The learning hypothesis. *Journal of Marketing Research*, 6(4):430–433.

Blake, T., Nosko, C., and Tadelis, S. (2015). Consumer heterogeneity and paid search effectiveness: A large-scale field experiment. *Econometrica*, 83(1):155–174.

Branco, F., Sun, M., and Villas-Boas, J. M. (2012). Optimal search for product information. *Management Science*, 58(11):2037–2056.

Branco, F., Sun, M., and Villas-Boas, J. M. (2016). Too much information? Information provision and search costs. *Marketing Science*, 35(4):605–618.

Bronnenberg, B. J., Kim, J. B., and Mela, C. F. (2016). Zooming in on choice: How do consumers search for cameras online? *Marketing Science*, 35(5):693–712.

Burguet, R. and Petrikaite, V. (2017). Targeted advertising and costly consumer search. Working paper.

Butters, G. R. (1978). Equilibrium distributions of sales and advertising prices. *The Review of Economic Studies*, 44(3):465–419.

Chan, T. Y., Wu, C., and Xie, Y. (2011). Measuring the lifetime value of customers acquired from google search advertising. *Marketing Science*, 30(5):837–850.

Chen, Y. and Yao, S. (2017). Sequential search with refinement: Model and application with click-stream data. *Management Science*, 63(12):4345–4365.

De Corniere, A. (2016). Search advertising. *American Economic Journal: Microeconomics*, 8(3):156–88.

De los Santos, B., Hortaçsu, A., and Wildenbeest, M. R. (2012). Testing models of consumer search using data on web browsing and purchasing behavior. *American Economic Review*, 102(6):2955–80.

De los Santos, B. and Koulayev, S. (2017). Optimizing click-through in online rankings with endogenous search refinement. *Marketing Science*, 36(4):542–564.

Du, R. Y., Xu, L., and Wilbur, K. C. (2019). Immediate responses of online brand search and price search to tv ads. *Journal of Marketing*, 83(4):81–100.

Dukes, A. and Liu, L. (2016). Online shopping intermediaries: The strategic design of search environments. *Management Science*, 62(4):1064–1077.

Duncan, C. P. and Olshavsky, R. W. (1982). External search: The role of consumer beliefs. *Journal of Marketing Research*, 19(1):32–43.

Fershtman, D. and Pavan, A. (2020). Sequential learning with endogenous consideration sets. Working Paper.

Fong, N. M. (2017). How targeting affects customer search: A field experiment. *Management Science*, 63(7):2353–2364.

Furse, D. H., Punj, G. N., and Stewart, D. W. (1984). A typology of individual search strategies among purchasers of new automobiles. *Journal of Consumer Research*, 10(4):417–431.

Gardete, P. and Hunter, M. (2020). Guiding consumers through lemons and peaches: An analysis of the effects of search design activities. Working Paper.

Ghose, A. and Todri-Adamopoulos, V. (2016). Toward a digital attribution model: Measuring the impact of display advertising on online consumer behavior. *MIS Quarterly*, 40(4):889–910.

Goeree, M. S. (2008). Limited information and advertising in the US personal computer industry. *Econometrica*, 76(5):1017–1074.

Golden, J. M. and Horton, J. J. (2020). The effects of search advertising on competitors: An experiment before a merger. *Management Science*, 67(1):342–362.

Goldfarb, A. and Tucker, C. (2011). Search engine advertising: Channel substitution when pricing ads to context. *Management Science*, 57(3):458–470.

Gossner, O., Steiner, J., and Stewart, C. (2020). Attention please! Forthcoming, *Econometrica*.

Greminger, R. (2020). Optimal search and discovery. Working Paper.

Hastings, J., Hortaçsu, A., and Syverson, C. (2017). Sales force and competition in financial product markets: The case of Mexico's social security privatization. *Econometrica*, 85(6):1723–1761.

Hong, H. and Shum, M. (2006). Using price distributions to estimate search costs. *The RAND Journal of Economics*, 37(2):257–275.

Honka, E. (2014). Quantifying search and switching costs in the US auto insurance industry. *The RAND Journal of Economics*, 45(4):847–884.

Honka, E. and Chintagunta, P. (2017). Simultaneous or sequential? Search strategies in the US auto insurance industry. *Marketing Science*, 36(1):21–42.

Honka, E., Hortaçsu, A., and Vitorino, M. A. (2017). Advertising, consumer awareness and choice: Evidence from the US banking industry. *The RAND Journal of Economics*, 48(3):611–646.

Iyer, G., Soberman, D., and Villas-Boas, J. M. (2005). The targeting of advertising. *Marketing Science*, 24(3):461–476.

Jeziorski, P. and Moorthy, S. (2017). Advertiser prominence effects in search advertising. *Management Science*, 64(3):1365–1383.

Joo, M., Wilbur, K. C., Cowgill, B., and Zhu, Y. (2013). Television advertising and online search. *Management Science*, 60(1):56–73.

Joo, M., Wilbur, K. C., and Zhu, Y. (2016). Effects of TV advertising on keyword search. *International Journal of Research in Marketing*, 33(3):508–523.

Katona, G. and Mueller, E. (1955). A study of purchase decisions. In *Consumer Behavior: The Dynamics of Consumer Reaction*, volume 1, pages 30–87. New York: New York University Press, 1955.

Ke, T. T., Shen, Z.-J. M., and Villas-Boas, J. M. (2016). Search for information on multiple products. *Management Science*, 62(12):3576–3603.

Ke, T. T. and Villas-Boas, J. M. (2019). Optimal learning before choice. *Journal of Economic Theory*, 180:383–437.

Kim, J. B., Albuquerque, P., and Bronnenberg, B. J. (2010). Online demand under limited consumer search. *Marketing Science*, 29(6):1001–1023.

Kim, J. B., Albuquerque, P., and Bronnenberg, B. J. (2017). The probit choice model under sequential search with an application to online retailing. *Management Science*, 63(11):3911–3929.

Kireyev, P., Pauwels, K., and Gupta, S. (2016). Do display ads influence search? Attribution and dynamics in online advertising. *International Journal of Research in Marketing*, 33(3):475–490.

Koulayev, S. (2014). Search for differentiated products: Identification and estimation. *The RAND Journal of Economics*, 45(3):553–575.

Li, H. and Kannan, P. (2014). Attributing conversions in a multichannel online marketing environment: An empirical model and a field experiment. *Journal of Marketing Research*, 51(1):40–56.

Ma, L. (2016). Only the interested learn: A model of proactive learning of product reviews. Working Paper.

Mayzlin, D. and Shin, J. (2011). Uninformative advertising as an invitation to search. *Marketing Science*, 30(4):666–685.

McFadden, D. (1989). A method of simulating moments for estimation of discrete response models without numerical integration. *Econometrica*, 57(5).

Moe, W. W. (2003). Buying, searching, or browsing: Differentiating between online shoppers using in-store navigational clickstream. *Journal of Consumer Psychology*, 13(1-2):29–39.

Moraga-González, J. L., Sándor, Z., and Wildenbeest, M. R. (2015). Consumer search and prices in the automobile market. Working Paper.

Moraga-González, J. L. and Wildenbeest, M. R. (2008). Maximum likelihood estimation of search costs. *European Economic Review*, 52(5):820–848.

Morozov, I. (2020). Measuring benefits from new products in markets with information frictions. Working paper.

Narayanan, S. and Kalyanam, K. (2015). Position effects in search advertising and their moderators: A regression discontinuity approach. *Marketing Science*, 34(4):612–627.

Newman, J. W. and Lockeman, B. D. (1975). Measuring prepurchase information seeking. *Journal of Consumer Research*, 2(3):216–222.

Newman, J. W. and Staelin, R. (1972). Prepurchase information seeking for new cars and major household appliances. *Journal of Marketing Research*, 9(3):249–257.

Rao, J. M. and Simonov, A. (2018). Firms' Reactions to Public Information on Business Practices: Case of Search Advertising. *Quantitative Marketing and Economics*, Forthcoming.

Ratchford, B. T. (2008). Consumer search behavior and its effect on markets. *Foundations and Trends® in Marketing*, 3(1):1–74.

Renault, R. (2016). Passive internet search. Working paper.

Robert, J. and Stahl, D. O. (1993). Informative price advertising in a sequential search model. *Econometrica: Journal of the Econometric Society*, pages 657–686.

Rothschild, M. (1974). A two-armed bandit theory of market pricing. *Journal of Economic Theory*, 9(2):185–202.

Rutz, O. J. and Bucklin, R. E. (2012). Does banner advertising affect browsing for brands? clickstream choice model says yes, for some. *Quantitative Marketing and Economics*, 10(2):231–257.

Rutz, O. J., Trusov, M., and Bucklin, R. E. (2011). Modeling indirect effects of paid search advertising: Which keywords lead to more future visits? *Marketing Science*, 30(4):646–665.

Sahni, N. and Zhang, C. (2020). Are consumers averse to sponsored messages? The role of search advertising in information discovery. Working Paper.

Seiler, S. (2013). The impact of search costs on consumer behavior: A dynamic approach. *Quantitative Marketing and Economics*, 11(2):155–203.

Shim, S. and Drake, M. F. (1989). Information search in the purchase of employment apparel: A synthesis of two theories. *Clothing and Textiles Research Journal*, 7(3):40–46.

Simonov, A. and Hill, S. (2021). Competitive advertising on brand search: Traffic stealing and click quality. *Marketing Science*, Forthcoming.

Simonov, A., Nosko, C., and Rao, J. M. (2018). Competition and crowd-out for brand keywords in sponsored search. *Marketing Science*, 37(2):200–215.

Stigler, G. J. (1961). The economics of information. *Journal of Political Economy*, 69(3):213–225.

Terui, N., Ban, M., and Allenby, G. M. (2011). The effect of media advertising on brand consideration and choice. *Marketing Science*, 30(1):74–91.

Tsai, Y.-L. and Honka, E. (2018). Non-informational advertising informing consumers: How advertising affects consumers' decision-making in the US auto insurance industry. Working paper.

Ursu, R., Wang, Q., and Chintagunta, P. K. (2020a). Search duration. *Marketing Science*, 39(5):849–871.

Ursu, R., Zhang, Q., and Honka, E. (2020b). Search gaps. Working Paper.

Ursu, R. M. (2018). The power of rankings: Quantifying the effect of rankings on online consumer search and purchase decisions. *Marketing Science*, 37(4):530–552.

Weitzman, M. L. (1979). Optimal search for the best alternative. *Econometrica*, pages 641–654.

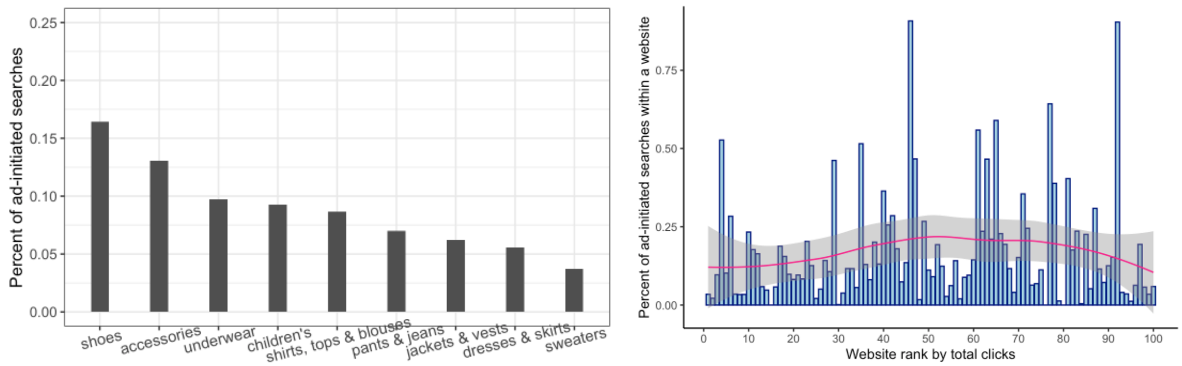
Whittle, P. (1981). Arm-acquiring bandits. *The Annals of Probability*, 9(2):284–292.

Wolinsky, A. (1986). True monopolistic competition as a result of imperfect information. *The Quarterly Journal of Economics*, 101(3):493–511.

Yang, S. and Ghose, A. (2010). Analyzing the relationship between organic and sponsored search advertising: Positive, negative, or zero interdependence? *Marketing Science*, 29(4):602–623.

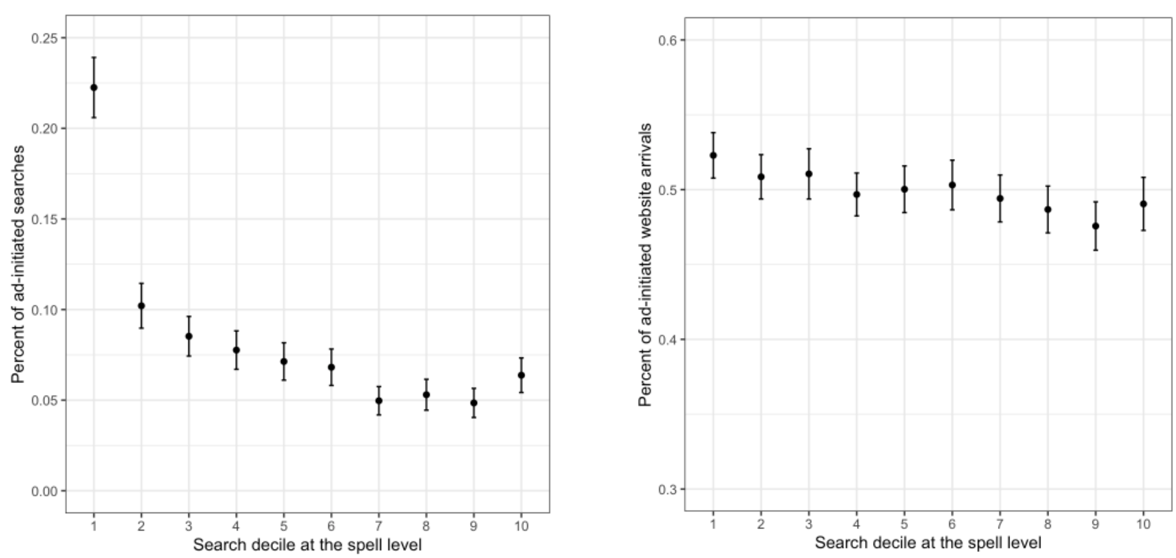
Figures and Tables

Figure 1: Percent of Ad-initiated Searches Across Subcategories and Websites



(a) Subcategories **(b) Websites**

Figure 2: Percent of Ads by Progress in the Spell



(a) Ad-Initiated Searches **(b) Ad-Initiated Website Arrivals**

Figure 3: Ad Types and their Search Frequency

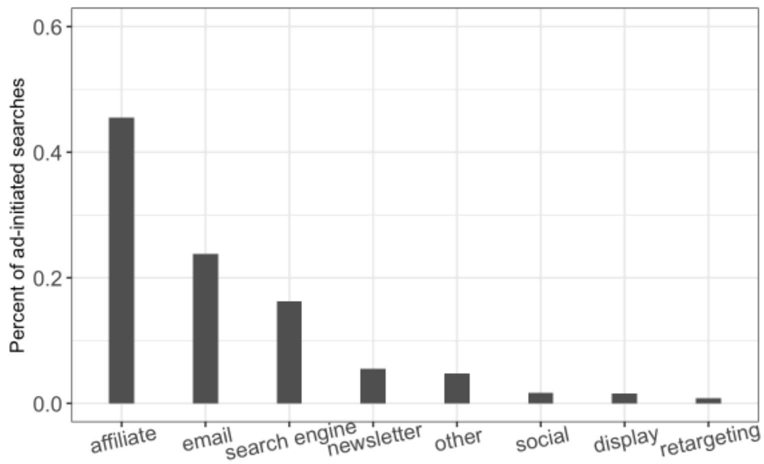
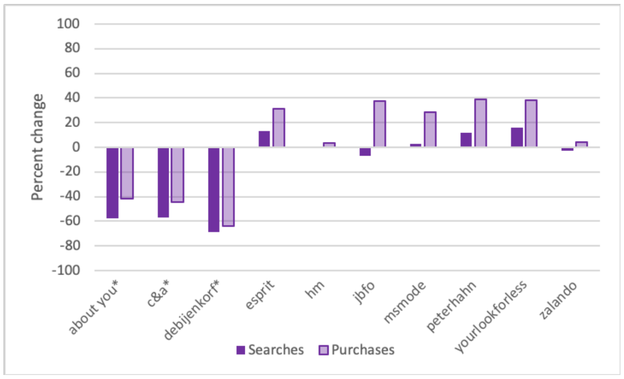


Figure 4: Percent of Website Landing Pages Across Ad Types

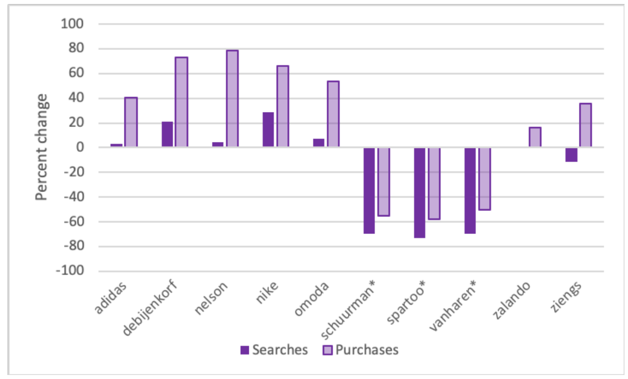
Table 1: Characteristics of Fashion Searches in a Spell

	Mean	Median	St. Dev.
Number of Clicks	75.72	32.00	2447.00
Number of Searched Websites	6.36	3.00	137.00
Duration (in minutes)	40.20	17.63	1162.05
Number of Products	22.88	9.00	596.00

Figure 5: Detailed Simulation Results



Subcategory 1

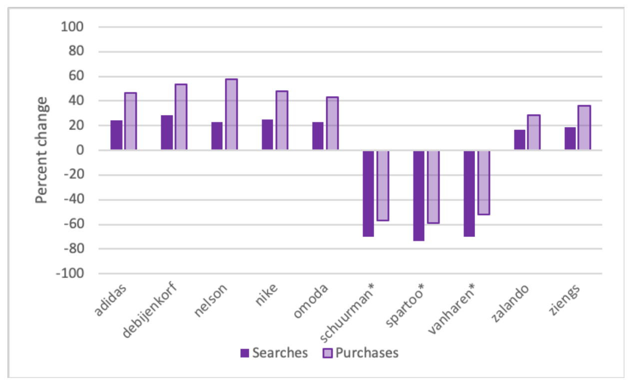


Subcategory 2

(a) Scenario 1 – No Advertising[a]

[a]Notes: Websites with an asterisk represent the three largest advertisers in a subcategory.

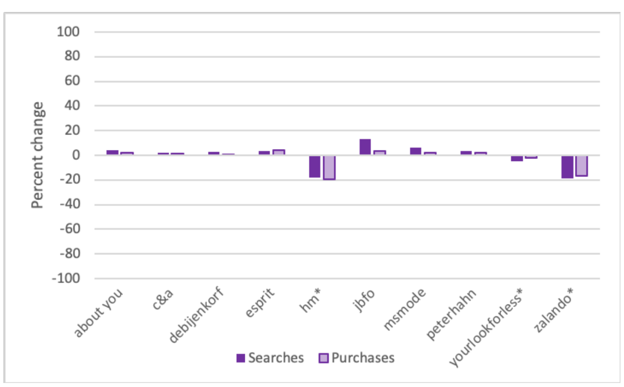
Subcategory 1



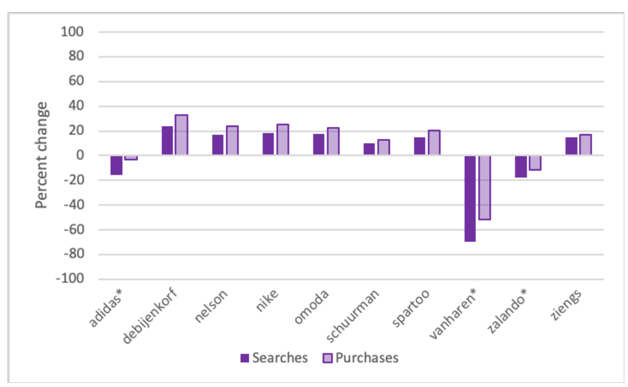
Subcategory 2

(b) Scenario 2 – No Advertising by the Largest Advertisers[a]

[a]Notes: Websites with an asterisk represent the three largest advertisers in a subcategory.



Subcategory 1



Subcategory 2

(c) Scenario 3 – No Advertising by the Websites with the Most Purchases[a]

[a]Notes: Websites with an asterisk represent the three websites with the most purchases in a subcategory.

Table 2: Website-level Summary Statistics Based on the Website Arrival Type

	Organic	Ad-Initiated	*T-stat*
Percent of Website Searches	0.45	0.53	–
Percent of All Transactions	0.71	0.29	–
Number of Clicks	8.37	4.10	30.74
Time Spent on Website (min)	4.32	2.28	31.75
Number of Searched Product Categories	0.99	0.66	44.64
Number of Searched Products	1.28	0.58	24.60
Standardized Price of Clicked Products	-0.14	0.16	-1.95

Notes: The last column reports the t-statistic for the difference in means.

Table 3: Average Statistics by Quantiles of Ad-Initiated Searches per Spell

	<20%ile	20-40%ile	40-60%ile	60-80%ile	>80%ile
Percent of Ads	0.00	0.04	0.13	0.36	0.85
Number of Clicks	56.26	153.20	91.26	52.60	24.19
Total Time Spent Searching (min)	28.09	76.89	50.79	32.19	12.57
Number of Websites	1.99	6.72	7.32	8.89	6.92
Number of Searched Product Categories	2.94	4.67	3.76	2.99	2.03
Number of Searched Products	9.97	23.38	13.18	5.17	0.80
Percent of Spells w/ a Transaction	0.31	0.43	0.30	0.12	0.017

Table 4: Example Illustrating Differences Across Models

			(i)	(ii)	(iii)
option	searched	ad	*Weitzman model*	*AP-strong model*	*AP-weak model*
1	1	0	$z1 \geq z2, z1 \geq z3, z1 \geq z4, z1 \geq z5$	$\cancel{z1 \geq z2}, z1 \geq z3, \cancel{z1 \geq z4}, z1 \geq z5$	$\cancel{z1 \geq z2}, z1 \geq z3, \cancel{z1 \geq z4}, z1 \geq z5$
2	1	1	$z2 \geq u1, z2 \geq z3, z2 \geq z4, z2 \geq z5$	$z2 \geq u1, \cancel{z2 \geq z3}, \cancel{z2 \geq z4}, \cancel{z2 \geq z5}$	$z2 \geq u1, z2 \geq z3, \cancel{z2 \geq z4}, z2 \geq z5$
3	1	0	$z3 \geq max\{u1, u2\}, z3 \geq z4, z3 \geq z5$	$z3 \geq max\{u1, u2\}, \cancel{z3 \geq z4}, z3 \geq z5$	$z3 \geq max\{u1, u2\}, \cancel{z3 \geq z4}, z3 \geq z5$
4	0	1	$z4 < max\{u1, u2, u3\}$	$z4 < max\{u1, u2, u3\}$	$z4 < max\{u1, u2, u3\}$
5	0	0	$z5 < max\{u1, u2, u3\}$	$z5 < max\{u1, u2, u3\}$	$z5 \leq max\{u1, u2, u3\}$

Table 5: **Monte Carlo Simulation Results**

	(i)	(ii)		(iii)	
Data Generating Model:		*AP-strong*		*AP-weak*	
Estimation Model:		*AP-strong*	*Weitzman*	*AP-weak*	*Weitzman*
	True values	Estimates (SD)		Estimates (SD)	
Utility					
Outside option	0.5	0.48	0.43	0.49	0.45
		(0.08)	(0.08)	(0.08)	(0.07)
Website 1	-1	-0.88	-0.87	-0.90	-0.88
		(0.06)	(0.08)	(0.07)	(0.08)
Website 2 (advertiser)	-0.5	**-0.49**	**-0.65**	**-0.43**	**-0.63**
		(0.07)	(0.07)	(0.07)	(0.07)
Website 3	-0.3	-0.27	-0.27	-0.27	-0.26
		(0.06)	(0.07)	(0.06)	(0.07)
Search cost (exp)					
Constant	-3	-2.97	-3.03	-2.86	-2.92
		(0.10)	(0.10)	(0.11)	(0.10)
Log-likelihood		-3,745	-4,071	-3,726	-3,995
Number of Observations		5,000	5,000	5,000	5,000

Table 6: Estimation Results

	Subcat. 1: "Shirts, tops, & blouses"					Subcat. 2: "Shoes"			
	(i)	(ii)	(iii)	(iv)		(i)	(ii)	(iii)	(iv)
	AP-strong	*AP-weak*	*Weitzman*	*Weitzman*		*AP-strong*	*AP-weak*	*Weitzman*	*Weitzman*
Utility					*Utility*				
aboutyou.com	-1.28***	-1.10***	-1.18***	-1.44***	adidas.com	-0.95***	-1.09***	-1.04***	-0.88***
	(0.04)	(0.03)	(0.03)	(0.03)		(0.03)	(0.03)	(0.02)	(0.04)
c-and-a.com	-0.78***	-0.61***	-0.68***	-0.90***	debijenkorf.nl	-1.64***	-1.81***	-1.75***	-1.52***
	(0.03)	(0.03)	(0.03)	(0.03)		(0.05)	(0.04)	(0.03)	(0.05)
debijenkorf.nl	-1.62***	-1.51***	-1.57***	-1.76***	nelson.nl	-1.33***	-1.42***	-1.33***	-1.18***
	(0.04)	(0.03)	(0.04)	(0.04)		(0.03)	(0.03)	(0.03)	(0.03)
esprit.nl	-1.66***	-1.71***	-1.72***	-1.62***	nike.com	-1.07***	-1.31***	-1.21***	-0.93***
	(0.05)	(0.04)	(0.04)	(0.04)		(0.03)	(0.03)	(0.03)	(0.03)
hm.com	-1.19***	-1.21***	-1.23***	-1.17***	omoda.nl	-1.40***	-1.53***	-1.45***	-1.23***
	(0.03)	(0.03)	(0.03)	(0.03)		(0.04)	(0.03)	(0.03)	(0.04)
jbfo.nl	-2.47***	-2.53***	-2.54***	-2.46***	**schuurman-shoenen.nl**	-0.64***	-0.64***	-0.78***	-1.05***
	(0.12)	(0.12)	(0.13)	(0.11)		(0.03)	(0.02)	(0.02)	(0.02)
msmode.nl	-1.69***	-1.72***	-1.72***	-1.62***	**spartoo.nl**	-1.09***	-1.05***	-1.14***	-1.52***
	(0.04)	(0.04)	(0.04)	(0.05)		(0.03)	(0.03)	(0.03)	(0.03)
peterhahn.nl	-1.74***	-1.83***	-1.85***	-1.76***	**vanharen.nl**	-0.79***	-0.79***	-0.87***	-1.13***
	(0.06)	(0.05)	(0.07)	(0.06)		(0.03)	(0.02)	(0.02)	(0.02)
your-look-for-less.nl	-1.39***	-1.42***	-1.45***	-1.37***	zalando.nl	-0.57***	-0.74***	-0.73***	-0.56***
	(0.04)	(0.04)	(0.04)	(0.04)		(0.03)	(0.02)	(0.02)	(0.03)
zalando.nl	-1.09***	-1.09***	-1.10***	-1.04***	ziengs.nl	-1.58***	-1.62***	-1.54***	-1.41***
	(0.04)	(0.04)	(0.03)	(0.03)		(0.05)	(0.04)	(0.03)	(0.04)
Number of previous website visits	0.20***	0.19***	0.16***	0.16***	Number of previous website visits	0.21***	0.21***	0.21***	0.20***
	(0.01)	(0.01)	(0.01)	(0.01)		(0.01)	(0.01)	(0.01)	(0.02)
Visit to a price discount page	1.94***	2.00***	1.76***	1.62***	Visit to a price discount page	1.47***	1.48***	1.18***	0.96***
	(0.03)	(0.03)	(0.04)	(0.04)		(0.05)	(0.04)	(0.03)	(0.04)
Outside option	2.07***	2.10***	1.98***	1.95***	Outside option	2.35***	2.25***	2.09***	2.11***
	(0.03)	(0.04)	(0.03)	(0.03)		(0.03)	(0.03)	(0.02)	(0.04)
Search cost (exp)					*Search cost (exp)*				
Constant	-3.97***	-4.18***	-4.15***	-3.71***	Constant	-5.13***	-5.28***	-4.97***	-4.11***
	(0.04)	(0.06)	(0.07)	(0.05)		(0.05)	(0.05)	(0.05)	(0.08)
Advertising				-2.16***	Advertising				-2.78***
				(0.08)					(0.07)
Observations	32422	32422	32422	32422	Observations	34812	34812	34812	34812
LL	-8974	-10381	-11546	-11095	LL	-13129	-15907	-18585	-17312
MAD (purchase)	34.59	36.77	41.13	36.55	MAD (purchase)	36.57	40.52	48.54	44.65
MAD (search)	59.15	65.73	93.86	74.84	MAD (search)	109.27	92.52	180.38	376.04
RMSE (purchase)	59.80	61.76	70.73	67.77	RMSE (purchase)	57.60	65.96	76.24	69.18
RMSE (search)	65.00	75.02	101.86	86.15	RMSE (search)	118.10	102.96	197.40	435.22

Standard errors in parentheses.

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

Notes: Websites in bold identify the three largest advertisers in each subcategory.

Table 7: Estimation Results (continued)

	Subcat. 3: "Pants & Jeans"					Subcat. 4: "Underwear"			
	(i)	(ii)	(iii)	(iv)		(i)	(ii)	(iii)	(iv)
	AP-strong	*AP-weak*	*Weitzman*	*Weitzman*		*AP-strong*	*AP-weak*	*Weitzman*	*Weitzman*
Utility					*Utility*				
c-and-a.com	-0.77***	-0.83***	-0.82***	-0.81***	asos.nl	-1.78***	-1.91***	-1.87***	-1.77***
	(0.04)	(0.03)	(0.03)	(0.04)		(0.07)	(0.06)	(0.07)	(0.06)
debijenkorf.nl	-1.35***	-1.42***	-1.41***	-1.40***	debijenkorf.nl	-1.46***	-1.57***	-1.56***	-1.47***
	(0.05)	(0.05)	(0.04)	(0.05)		(0.06)	(0.05)	(0.05)	(0.05)
esprit.nl	-1.33***	-1.39***	-1.36***	-1.32***	happysocks.nl	-1.74***	-1.76***	-1.74***	-1.69***
	(0.05)	(0.05)	(0.05)	(0.05)		(0.07)	(0.06)	(0.06)	(0.06)
g-star.com	-1.81***	-1.86***	-1.85***	-1.84***	hm.com	-1.12***	-1.18***	-1.17***	-1.11***
	(0.08)	(0.07)	(0.07)	(0.07)		(0.05)	(0.04)	(0.05)	(0.04)
hm.com	-0.86***	-0.92***	-0.91***	-0.89***	**hunkemoller.nl**	-0.65***	-0.42***	-0.49***	-0.65***
	(0.04)	(0.04)	(0.03)	(0.04)		(0.05)	(0.04)	(0.04)	(0.03)
jeanscentre.nl	-1.48***	-1.41***	-1.44***	-1.67***	**livera.nl**	-1.41***	-1.34***	-1.42***	-1.55***
	(0.07)	(0.05)	(0.05)	(0.05)		(0.06)	(0.05)	(0.05)	(0.05)
missetam.nl	-0.91***	-0.43***	-0.43***	-0.66***	mona-mode.nl	-2.04***	-2.19***	-2.13***	-2.02***
	(0.04)	(0.03)	(0.03)	(0.03)		(0.12)	(0.11)	(0.10)	(0.09)
tommy.com	-1.98***	-1.95***	-1.93***	-1.96***	ullapopken.nl	-1.45***	-1.55***	-1.52***	-1.42***
	(0.11)	(0.08)	(0.08)	(0.08)		(0.06)	(0.05)	(0.05)	(0.05)
your-look-for-less.nl	-1.21***	-1.28***	-1.27***	-1.24***	wibra.eu	-1.57***	-1.67***	-1.66***	-1.57***
	(0.05)	(0.04)	(0.04)	(0.05)		(0.06)	(0.05)	(0.06)	(0.06)
zalando.nl	-0.61***	-0.61***	-0.66***	-0.79***	**zalando.nl**	-1.06***	-1.05***	-1.09***	-1.18***
	(0.03)	(0.03)	(0.03)	(0.03)		(0.05)	(0.05)	(0.04)	(0.04)
Number of previous	0.17***	0.17***	0.16***	0.16***	Number of previous	0.12***	0.12***	0.10***	0.09***
website visits	(0.01)	(0.01)	(0.01)	(0.01)	website visits	(0.01)	(0.01)	(0.01)	(0.01)
Visit to a	1.78***	1.68***	1.57***	1.51***	Visit to a	1.83***	1.77***	1.79***	1.77***
price discount page	(0.05)	(0.05)	(0.05)	(0.05)	price discount page	(0.09)	(0.07)	(0.08)	(0.14)
Outside option	2.26***	2.23***	2.20***	2.16***	Outside option	2.21***	2.20***	2.13***	2.11***
	(0.03)	(0.03)	(0.03)	(0.03)		(0.05)	(0.04)	(0.03)	(0.04)
Search cost (exp)					*Search cost (exp)*				
Constant	-3.70***	-3.80***	-3.85***	-3.58***	Constant	-3.84***	-4.07***	-4.10***	-3.69***
	(0.06)	(0.05)	(0.04)	(0.06)		(0.06)	(0.07)	(0.06)	(0.08)
Advertising				-2.04***	Advertising				-1.96***
				(0.10)					(0.09)
Observations	27552	27552	27552	27552	Observations	17988	17988	17988	17988
LL	-7665	-8917	-9357	-9083	LL	-4466	-5294	-5681	-5497
MAD (purchase)	29.07	31.96	34.26	32.92	MAD (purchase)	14.32	15.32	18.45	17.74
MAD (search)	49.17	56.62	73.49	68.93	MAD (search)	24.05	25.71	49.99	44.63
RMSE (purchase)	59.25	61.71	64.55	64.91	RMSE (purchase)	25.92	27.05	32.57	32.39
RMSE (search)	52.31	80.95	93.85	78.36	RMSE (search)	26.14	36.14	57.28	49.13

Standard errors in parentheses.

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

Notes: Websites in bold identify the three largest advertisers in each subcategory.

Table 8: Simulation Results

	Subcat. 1	Subcat. 2	Subcat. 3	Subcat. 4
Scenario 1				
No advertising				
Total searches	-24.35%	-35.77%	-19.77%	-26.25%
Total purchases	-15.53%	-20.84%	-11.19%	-12.50%
Scenario 2				
No advertising by largest advertisers				
Total searches	-14.83%	-17.36%	-11.34%	-14.55%
Total purchases	-6.19%	-5.88%	-1.94%	-4.74%
Scenario 3				
No advertising by websites with the most purchases				
Total searches	-1.82%	-7.55%	-4.40%	-11.28%
Total purchases	-2.51%	-7.64%	-2.96%	-7.57%

10 Appendix

10.1 Data Cleaning

Our initial data source is the data used in Ursu et al. (2020b). We further cleaned the data on fashion searches, which initially consisted of 428,651 clicks, 40,735 sessions and 5,665 spells, as follows. We dropped sessions if they consisted of all non-search clicks (59 sessions), given that in this paper our focus is on active versus passive searches. Also, we dropped spells if the first session in a spell was dropped (i.e. if it contained only non-search clicks), since in such cases we may not observe the user's previous search activities (16 spells). Our final data sample consists of 427,768 fashion clicks with 40,625 sessions and 5,649 spells.

10.2 Classifying Ads by Type

We classified ad-initiated searches into eight types using the parameter component of a URL:

1. display – if the parameter component included keywords such as: display, banner, image.
2. email – if the parameter component included keywords such as: email, e-mail, mail, gmail, outlook, live.com.
3. newsletter – if the parameter component included keywords such as: nwl, newsletter, nieuwsbrief.
4. retargeting – if the parameter component included keywords such as: retarget, remarket.
5. search engine – if the parameter component included keywords such as: (a) gclid, gclsrc, dclid, or msclkid, (b) search engine names like google, bing, and yahoo, or (c) cpc, seo, ppc, sem, engine.
6. social – if the parameter component included keywords such as: social, instagram, facebook, fb, twitter.
7. affiliate – if the parameter component did not include (1)-(6) related variables but did include affiliate advertisers' names or affiliate ids such as: affiliate, refid, affid, partnerid, zanox, awin, daisycon, tradetracker, criteo, adtraxx, affilinet, copernica, and zanpid. Affiliate ads that contained information on the type of third-party website the ad was placed on (e.g. email, search engine, social, etc), were reclassified to match the groups identified above. Finally, based on

the click performed before an affiliate ad click, we classified affiliate ads placed on cashback websites (31%) as active searches. This is a conservative approach, since consumers may chose to search these cashback websites looking for fashion deals.

8. other – otherwise. For example, some URLs only included campaign ids, so we cannot identify the ad type.

10.3 Estimation Samples

We constructed four estimation samples corresponding to the four most commonly purchased product subcategories in our data: (1) "shirts, tops, and blouses", (2) "shoes", (3) "pants and jeans", and (4) "underwear". For each subcategory, we determined the top 10 most searched websites (accounting for approximately 65% of clicks in each subcategory), for which we estimated website intercepts. All other websites were grouped together into a composite website which we call "Other."

Since neither our AP models, nor the Weitzman model can accommodate revisits, we removed search revisits (to the a website) from the data, accounting for approximately 30% of observations. Also, we removed spells that had a search session within the last two days of our observation period, but no transaction, in order to avoid concerns about right truncation.[32] A small fraction (less than 1%) of spells contained more than one product purchased (after the changes we made to the samples), which we removed from the sample. In estimation, an ad-initiated search is a website where the arrival to the website (first click) was passive. The resulting estimation samples can be summarized as follows (Table A-1):[33]

Table A-1: Summary Statistics by Estimation Sample

	Subcat. 1	Subcat. 2	Subcat. 3	Subcat. 4
Observations	32,422	34,812	27,552	17,988
Spells	2,702	2,901	2,296	1,499
Converting Spells	359	316	271	152
Ads Searched	1,526	3,068	872	706
Ads Searched First/Ads Searched	0.63	0.50	0.80	0.78

In our data we only observe ads consumers clicked on. However, to more accurately capture the magnitude of the effect of passive search and to be able to consider an effect of advertising on

[32]Spells that end within the first week of our observation period (before February 23rd, 2018) were dropped from the original data sample, alleviating concerns about left truncation.

[33]Note that the reported number of observations includes an outside option for each spell.

search costs (the Weitzman model with advertising costs), we need understand the extent to which consumers might have been exposed to ads. We assume a consumer i was exposed to an ad from website j that she did not click if all of the following criteria are met: (i) consumer i clicked on an ad from website j in a different subcategory in the past, increasing her likelihood of receiving ads from the same website; (ii) consumer i had an open account with website j, increasing her likelihood of email and newsletter ads; (iii) website j advertises extensively in a given subcategory (more than the 90th percentile of the ad distribution in a subcategory), increasing the consumer's probability of being exposed to ads from this website; and (iv) consumer i clicked on at least one ad in the current spell, suggesting the consumer may be more likely to be exposed to ads (for example because she does not use ad blocker software; also, this allows us to be more conservative in our approach to infer ad exposure). For robustness, we also estimate our proposed model on the raw data, without any ads on websites that were not searched, and show that our results continue to hold. These results can be found in Table A-2 in Appendix 10.5.

In estimation, any ads that were not searched will be assumed to have occurred after the last searched website. To show that this assumption is in most cases innocuous, we use two approaches. First, we note that our robustness check, estimating the model on the raw data without any ads on websites that were not searched, also provides a robustness check for this assumption. Second, we demonstrate analytically in what narrow set of cases this assumption fails. If a consumer was exposed to ads she did not click on earlier in the search process than after the last searched website, then it means she searched other options after ad exposure. Let's denote by *ad* the ad the consumer did not click on, and by *next* any such options she searched after the ad she did not click on. In both AP models, if the consumer does not search an ad, then it must be that

$$z_{ad} < y, \tag{A1}$$

where y denotes the best option searched up to that moment in the search process. In contrast, because the consumer has searched an option after the ad she did not click on, then it must be that $z_{next} \geq y$. Thus, we conclude that

$$z_{next} \geq z_{ad}. \tag{A2}$$

Using this same logic for every website searched after the ad, we conclude that the ad the consumer did not click on has a lower reservation utility than all searched websites. This means that although the ad may have been shown earlier in the search, assuming it was presented to the consumer after all other searched websites will not produce a bias in the order of reservation utilities.

A bias may arise only because the reservation utility of the ad not clicked is compared against the utility of additional options when we assume it was presented to the consumer after all searched websites, rather than earlier. However, those additional options have reservation utilities z_{next} that are higher than z_{ad}, making it likely that their utilities are also higher than z_{ad} (since $z_j = u_j - \epsilon_j + fcn(c)$ – see Kim et al. (2010) for more details on the functional form of reservation utilities), thus not affecting the set of inequalities that identify our parameters. Only in the unlikely event that $fcn(c) - \epsilon_{next}$ is very large (e.g. very low search costs or very low utility shock draw), then our assumption would lead to a higher upper bound on reservation utilities z_{ad} than if we had observed ad exposure timing (which *may* lead to a higher reservation utility estimate, but does not need to). Given that we do not observe ad exposure timing, assuming consumers were exposed to ads they did not click on after the last searched website produces minimal (if any) bias in parameter estimates. Also, we note that our assumption is preferred over other alternatives, such as random or early exposure timing, because it does not disrupt the true order of reservation utilities, as demonstrated.

10.4 Estimation Procedure

The estimation procedure using the logit-smoothed AR simulator is standard in the literature (e.g.Honka and Chintagunta (2017), Ursu (2018)) and involves the following steps for the Weitzman (1979) model:

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 using the method proposed by Kim et al. (2010).

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

$$M^d = \sum_{k=1}^{4} e^{-v_k^d/\rho} \tag{A3}$$

and where ρ is a scaling parameter, chosen using Monte Carlo simulations. In our application, the scaling parameter equals $\rho = -15$.

5. The average of V^d over the D draws and over consumers and websites gives the simulated likelihood function.

It is straightforward to modify the above expressions for the AP-weak and the AP-strong models using the discussion in Section 6.2.

10.5 Additional Results

Figure A-1: Ad Types by Progress in the Spell

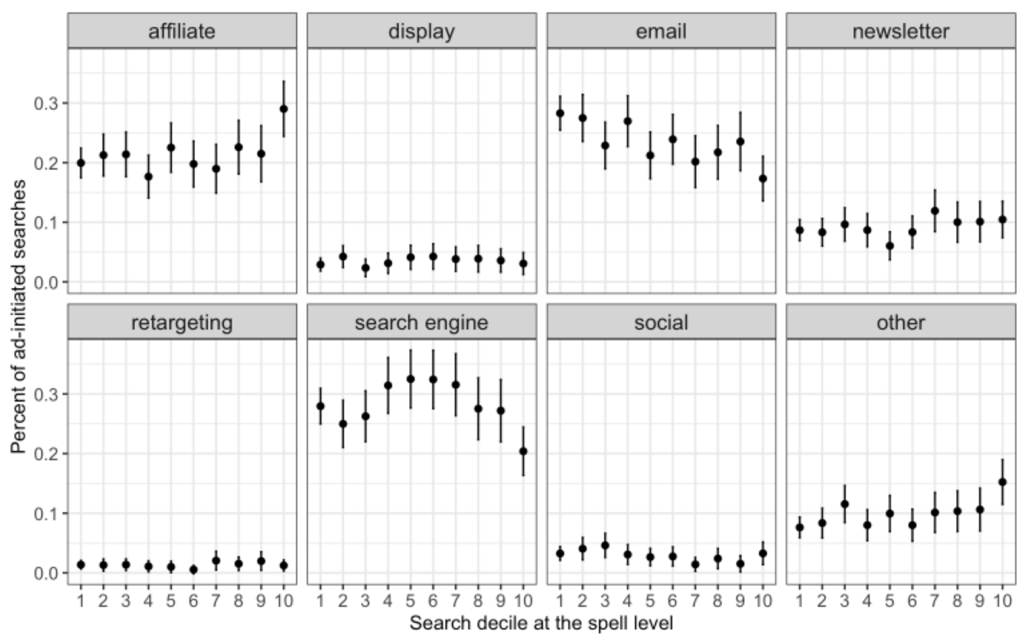


Table A-2: Robustness Checks

	(i)	(ii)		(iii)		(iv)		(v)
	AP-strong: search engine-active	*AP-strong: raw data*						
	Subcat. 1	Subcat. 1	Subcat. 2		Subcat. 3		Subcat. 4	
Utility			*Utility*		*Utility*		*Utility*	
aboutyou.com	-1.28***	-1.30	adidas.com	-0.96***	c-and-a.com	-0.77***	asos.nl	-1.78***
	(0.09)	(0.04)		(0.04)		(0.04)		(0.07)
c-and-a.com	-0.77***	-0.79***	debijenkorf.nl	-1.64***	debijenkorf.nl	-1.35***	debijenkorf.nl	-1.47***
	(0.07)	(0.03)		(0.04)		(0.06)		(0.05)
debijenkorf.nl	-1.64***	-1.64***	nelson.nl	-1.33***	esprit.nl	-1.33***	happysocks.nl	-1.73***
	(0.06)	(0.04)		(0.04)		(0.06)		(0.06)
esprit.nl	-1.67***	-1.65***	nike.com	-1.08***	g-star.com	-1.81***	hm.com	-1.12
	(0.09)	(0.05)		(0.03)		(0.08)		(0.04)
hm.com	-1.22***	-1.19***	omoda.nl	-1.40***	hm.com	-0.85***	**hunkemoller.nl**	-0.67***
	(0.06)	(0.03)		(0.04)		(0.05)		(0.04)
jbfo.nl	-2.45***	-2.44***	**schuurman-shoenen.nl**	-0.69***	**jeanscentre.nl**	-1.50***	**livera.nl**	-1.43***
	(0.18)	(0.13)		(0.03)		(0.07)		(0.06)
msmode.nl	-1.71***	-1.68***	**spartoo.nl**	-1.14***	**missetam.nl**	-0.92***	mona-mode.nl	-2.02***
	(0.14)	(0.04)		(0.04)		(0.04)		(0.10)
peterhahn.nl	-1.76***	-1.72***	**vanharen.nl**	-0.84***	tommy.com	-1.98***	ullapopken.nl	-1.46***
	(0.10)	(0.06)		(0.03)		(0.12)		(0.05)
your-look-for-less.nl	-1.41***	-1.39***	zalando.nl	-0.58***	your-look-for-less.nl	-1.21***	wibra.eu	-1.57***
	(0.05)	(0.04)		(0.03)		(0.05)		(0.06)
zalando.nl	-1.07***	-1.08***	ziengs.nl	-1.59***	**zalando.nl**	-0.63***	**zalando.nl**	-1.07***
	(0.07)	(0.03)		(0.04)		(0.04)		(0.05)
Number of previous website visits	0.20***	0.20***	Number of previous website visits	0.21***	Number of previous website visits	0.17***	Number of previous website visits	0.12***
	(0.01)	(0.01)		(0.02)		(0.01)		(0.01)
Visit to a price discount page	1.90***	1.94***	Visit to a price discount page	1.45***	Visit to a price discount page	1.76***	Visit to a price discount page	1.84***
	(0.07)	(0.04)		(0.05)		(0.06)		(0.08)
Outside option	2.06***	2.07***	Outside option	2.34***	Outside option	2.26***	Outside option	2.20***
	(0.04)	(0.04)		(0.04)		(0.04)		(0.05)
Search cost (exp)			*Search cost (exp)*		*Search cost (exp)*		*Search cost (exp)*	
Constant	-4.00***	-3.99***	Constant	-5.16***	Constant	-3.73***	Constant	-3.84***
	(0.11)	(0.06)		(0.07)		(0.07)		(0.07)
Observations	32422	32422	Observations	34812	Observations	27552	Observations	17988
LL	-9327	-9023	LL	-13242	LL	-7689	LL	-4490

Standard errors in parentheses.

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

Notes: Websites in bold identify the three largest advertisers in each subcategory.