Exploitation and Exploration: Improving Search Precision on E-commerce Platforms[1]

Wei Zhou
Eller College of Management, University of Arizona
weizhou1004@email.arizona.edu

Mingfeng Lin
Scheller College of Business, Georgia Institute of Technology
mingfeng.lin@scheller.gatech.edu

Mo Xiao
Eller College of Management, University of Arizona
mxiao@eller.arizona.edu

Zidong Wang
Alibaba Group
wangshang.wzd@alibaba-inc.com

(This draft: October 2020)

[1] We would like to thank Hidehiko Ichimura and Ashley Langer for their thoughtful comments. We are grateful to participants and/or reviewers at Symposium on Statistical Challenge in Electronic Commerce Research 2020, ISMS Marketing Science Conference 2020, OCIS Doctoral Consortium at Academy of Management 2020. We also thank Chunxiao Li, Zhe Yuan, Kaifu Zhang, and the economist team at the Alibaba Group for providing instrumenting help with data access. The views expressed herein are those of the authors and do not necessarily reflect the views of the Alibaba Group.

Exploitation and Exploration: Improving Search Precision on E-commerce Platforms

Abstract

E-commerce platforms match online buyers and sellers using their search technologies. Although a more precise search algorithm may improve search targetability, it may also reduce cross-selling opportunities as consumers spend less time exploring different products. We empirically quantify these tradeoffs through a collaboration with Alibaba Group. Specifically, we take advantage of a 2019 quasi-experiment on Taobao.com, in which the platform refined some product categories into finer subgroups in order to return more targeted search results to online shoppers. Using granular data on consumer search and purchase behaviors across multiple search sessions and product categories, we find that the improvement in search precision leads to a 37.3% increase in consumers' click-through rates and a 64.4% increase in gross merchandise volume in the product category we study. The improvement in matching outcomes in the short run, however, is accompanied with a substantial decrease in consumer engagement and unplanned purchases in the long run for consumers prone to spending longer time searching. On average, these consumers conduct 5.5% fewer searches, spend 4.1% less time on the platform, and decrease their spending on related categories by 2.2% in the following week after the search precision increases. Overall, our findings illustrate the tradeoff between exploitation and exploration in e-commerce search design that has not yet been previously documented in the literature.

Keywords: E-commerce, Consumer Search, Platform Design, Search Precision, Consumer Engagement

1. Introduction

Consumers arrive at an e-commerce platform with different keywords, sometimes “Fitbit Versa 2 special edition,” sometimes just “smart watch.” The platform’s search engine, keenly aware of the paucity and importance of search results page one, shoulders the responsibility of a traffic director. Should the platform just show Fitbit Versa 2 with different colors and add-ons? Or is Fitbit Charge 4 worth a mention? Or perhaps a big-name competitor, Garmin Forerunner 235? Or Xiaomi Mi Band, a relatively obscure, upcoming brand? In this example, the platform is trying to figure out the degree of search precision: precise search may deliver immediate efficacy, but too much precision may discourage consumers from exploring alternative options, new products, or relevant categories. Finding the sweet spot of balancing immediate gratification and long-term consumer engagement has spurred research interest in information systems, marketing, economics, and computer science(Ghose et al. 2012, De los Santos and Koulayev 2017, Yoganarasimhan 2018, Zhang et al. 2019), but so far our ability to push the research front forward is confounded by the endogeneity of observed search precision and constrained by the lack of detailed consumer search data.

Measuring the effects of improving search precision using observational e-commerce data has the following challenges. First, the precision of search results is endogenously determined by consumers’ unobserved preferences and their search actions such as sorting and filtering (Chen and Yao 2016, De los Santos and Koulayer 2017). Second, a typical search redesign goes beyond changing just consumers’ “consideration set,” as it may change the relative salience of different features of the search results, or elicit sellers’ strategic responses in product presentation, pricing, or other dimensions.[2] Third and more importantly, researchers rarely have access to individual-level search panel data that track each consumer across search sessions over time. Datasets examined in existing literature mostly contain a short episode of cross-sectional

[2] For example, Dinerstein et al (2018) studies a short-lived 2011 search redesign by eBay, in which consumers first identify an exact product, and then compare seller listing of that product, ranked mostly by price. They find that transaction prices fell by roughly 5 to 15% for many products, suggesting sellers engaged in more intense price competition during the redesign phase.

consumer search behaviors, abstracting away from the possibility of product discovery in the search process (Ghose et al 2019, Tan et al 2019).

To address these challenges, we take advantage of a detailed click-stream data set with quasi-experimental variation in the precision of search results. This data set comes from a particular change in search algorithms that refined the product categories on Taobao.com (Taobao henceforth), one of the world's largest e-commerce platforms.[3] Taobao creates multi-level categories to classify products so that its search engine can index and associate each product listed with different search queries. For example, "pet food" used to be a single category. Before the category refinement, consumers who submitted the query "cat food" would get pages with a mix of cat food and dog food in the search results. By refining the category "pet food" into two subcategories "cat food" and "dog food", the search engine is more likely to retrieve only cat food. As a result, the match between consumers' queries and sellers' relevant products was substantially improved after the category refinement. Since consumers were not aware of these behind-the-scenes adjustments of search algorithms, we can causally identify how consumers respond to the improvement in search precision and estimate the economic tradeoffs associated with it.

Our search dataset is at a very granular level, including around 7 million consumers over two-year period ending in December 2019. Two advantages of our dataset help overcome the limitations in previous studies on consumer search. First, our data provide a more comprehensive observation of consumers' engagement with the e-commerce platform. Most previous studies cannot connect searches made by the same user over time when using data from PC search. Consumers are not often signed in when searching on the browser/computer even if they have an account on the platform (Bake et al, 2016). We overcome this limitation by using consumer search data from mobile transactions, which account for over 95% of transactions made on Taobao. Consumers are automatically logged in when starting a search session on their mobile phones. We can track multiple sessions across time conducted by the same consumer.[4] For each consumer, we are able to link

[3] Taobao is an e-commerce platform that enables third-party sellers to sell products online. Unlike Amazon.com, Taobao does not sell directly to consumers. The innovation of search technology is at heart of its core business as to facilitate transactions between sellers and buyers. This also provides us an ideal environment to observe how buyers and sellers adapt to the advance of search technology.

[4] All consumers are anonymized in our sample. We do not use any personal information of consumers for the analysis.

their search sessions from the first query search all the way through to either a purchase or an abandonment of the search. We not only know how many clicks or purchases a consumer makes in a search session but also the total time she engages with the platform. This feature of our data allows us to go beyond the instantaneous effect within a search session to a much longer time horizon. Second, we focus on a basket of related products instead of a single category. Previous studies use search data in a single-category setting such as hotels and books (Ghose et al. 2012, Hong and Shum 2006). In contrast, sellers sell hundreds of thousands of categories of products on Taobao.com. Consumers often search across categories and refine search queries across search sessions. Later search queries are related to former search results, which generates cross-category dependencies (Liu and Toubia 2020). Our ability to track consumers across multiple categories allows us to look into the spillover effects across relevant product categories responding to the change in search precision.

Although the data seem to give us endless possibilities, we choose to focus on the product category "Trash Can" and its related categories of home and personal cleaning supplies such as mops, trash bags, vacuums, air purifiers, and storage racks.[5] We pick trash cans precisely because of its "unmarvelousness": it is a normal household item which has no easy substitute, it is a well-defined product, and it has steady demand throughout the year. Basically, it is a product we can reasonably argue as relatively immune to seasonality, stockpiling, trends and fads, or marketing gimmicks. On April 17, 2019, Taobao refined the product category "Trash Can" into two independent subcategories "General Trash Can" and "Smart Trash Can,"[6] and this category refinement event defines our treatment group as consumers whose search queries are related to "Smart Trash Can" or "Trash Can."[7] Our control groups come from several unaffected product categories that are for house cleaning purposes and share the parallel trends with trash cans in click-through rates and purchase rates before the category refinement. We use consumers whose search queries are related to "robot vacuum" as the control group in our main analysis and the others as robustness checks.[8] We first document the rise of search

[5] Section 7.2 explains how we define relevant categories to trash cans.

[6] Smart trash cans are a special type of trash cans which use advanced infrared sensor technology to open the lid automatically when consumers approach the bin, and close when consumers walk away.

[7] Taobao started to gradually rollout category refinement in selected categories from the start of 2019. In Table A1, we list other product categories that went through category refinement around April 2019.

[8] In section 6.4, we re-estimate the difference-in-differences model using the category "air purifier", "vacuum" and "mop" as the alternative control group respectively.

precision after the category refinement in the treatment group. We then employ a flexible difference-in-differences, and sometimes a triple difference design[9] to compare the changes in consumers' matching outcomes and search intensity in the treatment group before and after the category refinement relative to the same changes over time among consumers in the control group. More importantly, we track down consumers' activities on the platform in the following week after the initial search to measure the effect of the category refinement on consumer engagement and cross-selling in relevant categories.

We have four sets of main results, speaking to matching outcomes, search intensity within the same search session, consumer heterogeneity, and consumer engagement one week after the initial search. First, we find that the matching between consumers and products significantly increases with the improvement of search precision. Specifically, the average click-through rate of consumers who search for smart trash cans increases by 37.3% after the category refinement, relative to consumers who search for unaffected categories. The purchase rate of smart trash cans also increases by roughly 36.1%, resulting in a 64.4% increase of gross merchandise volume in the category of smart trash cans. These results are qualitatively and quantitatively robust to many alternative samples and alternative selections of control groups. We find little gain, however, for consumers who search general trash cans. We discover that this differential effect is driven by both Taobao's targeted allocation of search traffic and a surge of entry of smart trash can product listings right after the category refinement. Taobao gradually increased the ratio of search traffic to the "smart trash can" category from to 15% to 25% in the half-year span after the category refinement. In the meantime, the more targeted search traffic attracts more sellers to enter the category "Smart Trash Can," boosting the weekly number of smart trash can product listings on the platform by 150%. As smart trash cans are niche products that cater to a small population of consumers, consumers who search for smart trash cans would only get a few listings of smart trash cans among the multitudes of general trash cans before the category refinement, but the situation is reversed after the refinement. For consumers who search for general trash cans, their search results pages

[9] Consumers who search "smart trash can" and "trash can" experienced different treatment intensity, as we will explain in Section 5.2.

did not experience much of a change with the refinement. Therefore, the effect of the category refinement on matching outcomes is only demonstrated for consumers searching for smart trash cans.

Second, accompanying the increase in matching outcomes is a decrease in consumers' search intensity within a search session, measured by the number of listings a consumer views (by scrolling down her screen), the number of clicks she makes, and the total time she spends on these clicked listings. We show, again, that the effects are mostly for consumers who search for smart trash cans instead of general trash cans. Conditional on a consumer's click into any listing, the total number of viewed listings decreases by 4.4%, the number of clicked listings decreases by 3.7% and the total time spent viewing the clicked listings decreases by 6.6% after the category refinement.

Third, we show that the gain for consumers who search for smart trash cans also depends on consumer heterogeneity as indicated by their search queries. Specifically, we categorize two sub-groups of consumers: one group has specific shopping needs and knows exactly which products to buy, for example, "Smart Trash Can+ Xiaomi+ 7L + White," while the other group has a sub-category of products in mind, such as "sensor motion trash can", but is open to the arrays of options in this sub-category. We find that although both groups of consumers seem to benefit from higher search precision, they gain in different fashions. For consumers who have specific shopping needs, it is just an increase in matching outcomes without a reduction in search intensity. We suspect this group of consumers did not search much anyway before the category refinement. The real winners are consumers who only have a category of products in mind. On average, these consumers' click-through rate increases by 20.5% , their purchase rate increases by 31.9%, and they view 4.3% fewer listings, clicked into 3.7% fewer listings, and spend 6.2% less time on the clicked listings after the category refinement. This result is a clear indication of the rich consumer-level heterogeneity carried by their search queries, and a clear demonstration of how search design can better exploit this type of heterogeneity.

Combining the above results, it seems that Taobao's category refinement generates an overall efficient shopping experience for consumers who search for smart trash cans: a consumer announces her shopping intention, and the search engines respond with product listings precisely catering to her intention, and the consumer purchases quickly without much fuss or hassle. Does this gain in efficiency come at some cost? Is this the shopping environment that the e-commerce platform wants to create?

To answer these questions, we have to consider the fact that an e-commerce platform is a multi-product, multi-category (virtual) shopping mall. Demand-spillover and cross-selling, or achieving economies of scope, is essential for its business model (Basker et al 2012, Hwang and Park 2016, Rhode and Zhou 2019). To investigate the longer-term effect of search precision improvement, we follow the literature to divide consumers into "goal-directed" and "exploratory" types by the time they typically spent on the platform before category refinement.[10] The "goal-directed" type care about the efficiency of gathering information in the search process. The positive experiences of quickly getting what they want make this type of consumers more likely to return to the platform for shopping. But the "exploratory" searchers enjoy browsing and scanning various product categories on the platform without predefined shopping goals. For example, consumers may want to buy a rice cooker but does not know which particular product has the features she needs. After browsing some listings and trying different search queries, they may realize, instead of a rice cooker, what they need is actually a steamer.

Our fourth result shows that for "exploratory" consumers, the improvement in matching outcomes in the short run is accompanied with a substantial decrease in consumer engagement and purchases in related categories in the long run. Too precise search results may take away the fun of exploration, disincentivizing them from revisiting the platform to look through relevant products. For this type of consumers, we find a significant decrease in engagement with the platform in the following week when the search precision increases. The average number of days they visit the platform during a week decreases by 4.1%, the number of searches in a week decreases by 5.5%, and the number of search queries they use decreases by 5.6%. As a consequence, these consumers decrease their spending in the relevant product categories by about 2.2% in the following week when the search precision improves. As a comparison, although the "goal-directed" searchers seem to engage more with the platform and make more purchases in relevant categories, the magnitude of this gain is

[10] Previous studies have documented two shopping motives in consumer search behaviors: goal-directed and exploratory search (Janiszewski 1998, Wolfinbarger and Gilly 2001, Tam et al 2006, Chiou and Ting 2011, Pfeiffer et al 2020).

multi-folds smaller than the effects on the "exploratory" type. The loss in consumer engagement and cross-selling dominates the gain.

Taken together, our four sets of findings illustrate the tradeoff between exploitation and exploration in platform search design. A search engine, which direct search traffic, helps consumers to exploit and explore the search results. A more precise search engine results can generate immediate efficacy of satisfying consumers' shopping needs and boost the platform's transaction volume, but too much of precision can throttle stimulation and exploration, discouraging consumers from engaging with the platform. Although we cannot pinpoint the optimal level of search precision, our results clearly indicate that there is no single cut as the right amount of search precisions depends on product categories, the distribution of consumer preferences as well as their search habits, and the potential for scope economies.

To our awareness, we are the first empirical study that tracks consumer search over multiple search sessions and multiple product categories to demonstrate the multi-faceted effects of search precision improvement. Early in the literature, researchers show that grouping similar web pages into categories can improve the precision in the retrieved results and enhance user search experience (Xing et al. 2008, Bilal 2012). However, theories and experiments soon point out that more targeted and customized search results may discourage consumers from exploring new products, resulting in missed opportunities of cross-selling (Fong 2017, Hagiu and Wright 2020, Rhodes et al 2020). The exploitation-exploration dilemma is not unique to search design --- many research fields, ranging from organizational innovation, depression treatment, to reinforcement learning demonstrate similar conflicts (March 1991, Currie and MacLeod 2020, Sutton and Barto 1998). Guided by theory (Yang 2013, Zhong 2019) and enabled by our massive, granular consumer search data, we fill in the blank by illustrating these tradeoffs in the search engine design. Our results suggest that e-commerce platforms balance the short-term gains from increasing search precision with the long-term benefits from encouraging consumer exploration.

We will proceed by discussing our contribution to the literature in detail in Section 2. After describing our setting and data in detail in Sections 3 and 4, we present our empirical strategy in Section 5 and results from Section 6 to 8. Section 9 discusses interpretations, implications, and caveats of our results as concluding remarks.

2. Literature Review

Our study is related to three strands of literature: consumer search, platform search design, and the exploitation-exploration model. In this section, we briefly talk about the key models and results that inform the development of our paper and how our research contributes to the existing literature.

2.1 Consumer Search

This paper is related to the literature on consumer search. Starting from the seminal work of Stigler (1961) on the economics of information, theoretical studies have centered around how information frictions affect consumer welfare, market structure, and equilibrium outcomes (Wolinsky 1986, Stahl 1989, Anderson and Renault 1999). Consumer search models indicates that substantial search costs can discourage consumers from considering all available products in the market, thereby resulting in price dispersion (Varian 1980, Burdett and Judd 1983).Growing empirical studies have developed techniques to quantify the magnitudes and consequences of consumer search costs in various markets.[11] Early studies focus on searching in offline retail markets and assume that price is the main characteristic of the goods consumers are uncertain of when they are searching (Sorensen 2000, Hortacsu and Syverson 2004, Hong and Shum 2006). The rise of e-commerce draws the research attention to the more complex process of multi-attribute search as the detailed web browsing data become available (Bronnenberg et al 2016). Recent empirical studies examine consumers search behaviors within a platform when the purpose of search is for a good fit (Kim et al 2010, Koulayev 2014). There papers often consider a single category setting where consumers have an exact product in mind and their purpose of searching is to acquire information and resolve uncertainty about the product. A critical identification assumption is that consumers have rational expectation about the distribution of product attributes before searching (Honka et al 2019). [12] Thus a consumer stops searching either because of a high valuation for the

[11] For example, prescription drugs (Sorensen,2000), gasoline markets (Mitsukuni and Marc 2018, Luco 2019), supermarkets (Wildenbeest 2011), mutual fund (Hortaçsu and Syverson, 2004), automobile markets (Moraga-Gonzales et al 2018), personal consumer markets (Li et al 2017), mortgage markets (Alexandrov and Koulayev 2018), illicit drugs (Galenianos and Gavazza 2017)

[12] A further distinction of search models in the literature is the search method consumers are using when searching. Theoretical papers imply that consumers adopt either a sequential or simultaneous search model (McCall 1970, Weitzman

products already found which resulted in a successful search or because of a high search cost that discouraged consumers from continue searching.

Instead of examining the impacts of search costs, we contribute to the literature by empirically identifying search *quality* as another essential component of search frictions, especially in online markets. We join two recent theoretical papers in defining search quality as the precision of search results that can be altered by the search technologies of online retail platforms (Yang 2013, Zhong 2019).[13] Consistent with the theoretical prediction in Yang (2013), we confirm that a decrease in search costs and an increase in search quality have different qualitative effects. We find that an increase in search quality can discourage consumers from exploring unrelated products and decrease their engagement with the platform in the long run. Our findings thus challenge the conventional wisdom that minimizing search frictions is always the optimal strategy for e-commerce platforms (Brynjolfsson and Smith 2000).

By considering the role of search technologies in shaping the search results, our paper is also related to recent studies on the effects of ranking algorithms on consumer choices (Ghose et al 2012, 2014; Chen and Yao 2017; De los Santos and Koulayev 2017; Ursu 2018). We contribute to this literature by easing two assumptions in previous search models: (1) rational expectation and (2) category independence. First, we incorporate bounded rationality into the traditional search model. Consumers are not supposed to know exactly what to buy and expect correctly which products to encounter at the beginning of a search process. Second, we track consumer search activities across multiple categories. Further, we allow the precision of search results for one category to impact consumer search behaviors in another category. In contrast, most research considers consumer search behaviors in a single category. Table 1 compares our paper with recent papers on consumer online search behaviors.

[Table 1 about here]

1979, Burdett and Judd 1983, Stahl 1989). Researchers have developed empirical tests to differentiate these two search methods based on search path data (De Los Santos 2012, Honka and Chintagunta 2017).

[13] Yang (2013) incorporates the quality of search into the traditional search model to explain how the widespread of Internet leads to the long tail effects. He finds that a decrease in search costs and an increase in search quality have different qualitative effects. Zhong (2019) incorporates search precision into the search process. His model suggests that when the search precision is extremely high, increasing precision could discourage consumer search and lead to higher market prices.

2.2 Platform Search Design

Our study also contributes to the research on platform search design. E-commerce platforms match buyers and sellers using search technologies and recommendation systems (Bakos 2001, Brynjolfsson et al. 2011).[14] There has been a productive effort in optimizing search designs to increase consumer surplus and boost search engine revenues (Chen and Yao, 2017, Ghose et al, 2019, Gu and Wang 2018, Gardete and Hunter 2019, Zhang et al. 2019, Yoganarasimhan 2020). For example, Chen and Yao (2017) highlight the value of refinement tools in consumer online search and predict consumers will search less and have lower utilities when sorting or filter options are not available. Gu and Wang (2018) discuss how the optimal information layout of platform search design needs to consider consumers' cognitive costs when deciding what types of product attributes should be presented in the outer layer of search results. Ghose et al (2019) show that platforms can improve consumer online search experience by incorporating social textual content on the summary page of search results. However, most of these papers focus on (1) the short-run effects of platform designs on consumer search and purchase behaviors and (2) the changes from the demand side, abstracting away from the possibility of changes from the supply side in the long run that can affect market equilibrium outcomes.

Our long search panel data allow us to make two contributions to this literature. First, we analyze the effects of platform design on consumer behaviors both in the short run and in the long run. Our findings imply that search algorithms that increase search engine revenues in the short run may not be optimal in the long run. Second, we extend the analysis to the supply side and examine how sellers adapt to the changes in platform design in the long run. We document sellers' strategic entry behaviors after the search precision improves. The closest paper to ours is Dinerstein et al (2018), which highlights the trade-off of efficient platform design. They show that the platform design could not only reduce consumers' search costs, but also intensify price competition among sellers. In contrast, we focus on the role of platform in splitting the markets to improve matching between buyer and sellers. Our results indicate that niche products gain more market shares and the distribution of sales becomes flatter as search precision improves. In term of findings, our paper is also related

[14] By alerting the research attention to the matching role of e-commerce platforms, our study is also closely related to a growing number of studies on peer-to-peer service markets, where platforms play an active role in matching heterogeneous buyers and sellers (Hitsch et al 2010, Einav et al 2016, Chen and Shelton 2016, Fradkin 2017, Horton 2018).

to the literature on product design and long tail effect in online markets (Kuksov 2004, Brynjolfsson et al 2011, Bar-Issac et al 2012, Yang 2013, Larson 2013).

2.3 Exploitation and Exploration

We finally build on the exploitation and exploration models that have been well studied in the computer science, statistics, organization science, and economics literature (Schumpeter 1934, Holland 1975, Kuran 1988, March 1991, Sutton and Barto 1998, Benner and Tushman 2003). By refining and extending existing knowledge or solutions, exploitation can bring us great certainties of short-term gains. On the other hand, exploring new ideas and experimenting alternatives can help us increase long-term competencies and adapt to future environmental changes. Researchers have applied the theories of exploitation-exploration in the study of doctor drug prescription (Currie and MacLeod 2020), organizational innovation (Sorensen and Stuart 2000, Benner and Tushman 2003), professional hiring (Groysberg and Lee 2009), reinforcement learning (Holland 1975, Sutton and Barto 1998), and recommendation system design (McInerney et al 2018, Dzyabura and Hauser 2019). Although diverse in research topics, these studies can be unified in the multi-armed bandit framework where agents solve complex dynamic programming problems (Lai and Robbins 1985, Bubeck and Cesa-Bianchi 2012). The optimal decision rule hinges on balancing immediate rewards with the long-term benefits from learning about the distribution of rewards to inform future choices (Dzyabura and Hauser 2019).

Our study extends the literature on exploitation and exploration to the study of platform search design. The problem of determining the optimal precision of search results share the property of multi-armed bandit problems. Search engines can exploit short-term gains from immediate transactions by providing most relevant products to consumers. Alternatively, they can reap the long-term benefits from continuing consumer engagement and cross-category spillovers by displaying diverse and serendipitous items in the search results.[15]

[15] Studies on recommendation systems have argued that predictive accuracy should not be the only focus of recommendation algorithms (McNee and Konstan 2006, Fleder and Hosanagar 2009). Many authors suggest that platforms should avoid recommending products that are most likely to be chosen. Instead, recommendations should include diverse items that are not similar to each other and serendipitous items that are unexpected and relevant (Delgado-Battenfeld, and Jannach 2010, Zhou et al. 2010, Vargas and Castells 2011, Adamopoulos and Tuzhilin 2014).

To the best of our knowledge, our paper is among the first in the literature to identify the tradeoff between exploration and exploitation for search engine design.

3. Empirical Setting

3.1 Product Categorization and Search Engine Indexing

E-commerce platforms use multi-level categories to classify products into different categories so that search engines can index and associate each product with different search queries. Figure 1 provides an example of category hierarchy on e-commerce platforms. "Sports" is in the first-tier category, under which there are four second-tier categories: "Sports shoes", "Sport clothes", "Equipment", and "Accessories". The category "Sports Shoes" is further refined into three third-tier subcategories: "Tennis Shoes", "Soccer Shoes", and "Basketball Shoes".

[Figure 1 about here]

Categorization can not only directly navigate consumers to the products they want, but also organize all products for the search engine to retrieve. Consumers can directly go through the category menu and choose these product categories to search. Alternatively, they can start a search session by typing a specific query into the search box. Once the search query is understood, the search engine relates several product categories that are most likely to match the search query and further restricts its attention to the products that belong to these product categories.[16] Then the search engine ranks all these products and delivers search results to consumers. In this matching and ranking process, product categorization is crucial in helping the search engine understand consumers' shopping needs behind search queries (Xing et al. 2008, Bilal 2012).

Without opening physical stores, the most essential way for firms to "locate" themselves in this electronic world is to list their products into a particular category and let the search engine guide consumers toward them. Online search is the communication process between consumers and search engines where

[16] Search engines follow three primary steps to generate results from web documents: crawling, indexing, and ranking (Aggarwal 2018). As for e-commerce search engines, categorization plays a major role in indexing products. The indexing can determine the rankings orders of products in the search results.

consumers express their demands through search queries. Search engines are answer machines that try to pull out relevant web pages or products in the hopes of solving the searcher's query by indexing products for relevant keywords. In other words, if a product is indexed for a keyword, it will show as a result when a customer uses such search term in the search bar. The primary method of getting indexed for a keyword on e-commerce platforms is by selecting a relevant category. For example, sellers who list their products under the category "Tennis Shoes" are more likely to be indexed with keywords related to tennis shoes. Their products, however, are less likely to show up as a result for consumers who search for soccer shoes. Thus, the category where they list products can influence how they are found and what sellers they are competing against. Figure 2a is the snapshot of a seller's category selection process on Alibaba. The seller has to choose one specific category node to list a product. See Figure 2b for the similar process on Amazon.

[Figure 2 about here]

3.2 A Quasi-Experiment: Product Category Refinements on Alibaba

Alibaba is the leading e-commerce company in the world, with its three primary e-commerce sites – Taobao, Tmall and Alibaba.com – boasting 654 million active users annually and a gross merchandise volume (GMV) of $853 billion in 2018[17] and capturing a 55.9% share of all Chinese e-commerce retail sales[18]. Its sites mainly work as the marketplaces to connect various third-party sellers and willing buyers. Similar to eBay, Taobao is consumer-to-consumer focused, enabling small businesses and individuals to reach out consumers. In contrast, Tmall is a business-to-consumer type of marketplace focusing more on large companies and multinational brands such as Nike and Apple. While Taobao and Tmall cater individual consumers, Alibaba.com is a business-to-business trading platform for manufacturers and suppliers to find vendors and purchase merchandise in bulk.

To provide better matching outcomes and increase purchase conversion, Alibaba refined part of categories during 2019, generating quasi-experimental variation for our identification. See Table A1 in the appendix for a partial list of re-categorization. There are two types of category refinements. The first is to

---

[17] https://www.alibabagroup.com/en/news/press_pdf/p190515.pdf
[18] https://www.emarketer.com/content/alibaba-jd-com-lead-in-china-but-a-few-others-are-making-dents-too

separate a combined category into two independent categories. The products in these two categories are different from each other. Because of the very few sellers in both categories, they were once bundled together is to save administrative costs of managing and monitoring. For example, the e-cigarette and e-cigarette accessories both belonged to the same leaf category before the refinement. Then Alibaba divided this combined leaf category into two: an "E-cigarette" category and an "E-cigarette accessory" category. The other type of refinements is to single out one or several special types of products from a general product category. For example, the category "Tissue" was refined into three subcategories: "Paper Towel", "Napkin", and "General Tissue". By its nature, paper towel, napkin, and tissues are all paper products. However, they are used at different scenarios. These are all behind-the-scenes changes in search algorithms that consumers are not aware of.

After the category refinements, the search engine gives more targeted search traffic to the products belonging to the new product category. We use the refinement of the category "Trash Can" to show how it works.[19] In Figure 3, we can imagine each column of the right table to be what consumers actually see in their mobile phones. Before the category refinement, consumers who submit the query "Smart Trash Can" will see mostly general trash cans in the search results. After the category "Trash Can" was refined into two subcategories "Smart Trash Can" and "General Trash Can", consumers will find many smart trash cans in the search results

[Figure 3 about here]

4. Data

4.1 Sample Selection

To estimate the effect of the category refinement, we focus on the refinement of one particular category. After 4/18/2019, Alibaba divided the category "Trash Can" into two subcategories: "Smart Trash Can" and "General Trash Can". Smart trash cans are one special type of trash cans which use advanced infrared sensor technology to open the lid automatically when consumers approach the bin, and close when consumers walk away. Thus,

[19] In Section 7.1, we provide empirical support for this argument to show the underlying mechanism of category refinements.

consumers who want to buy a smart trash can may not be satisfied if they see general trash cans in the search results.

We choose this product category for two reasons. First, it has quite steady demand throughout the year. [20]Other product categories on the platform exhibit a time trend of demand, complicating our estimation of the effect of platform search design.[21] Second and more crucially, this category refinement brings us two natural treated groups with different treatment effects so that we can tease out potential selection bias in category refinements that threatens our identification. As a niche product that caters to a small population of consumers, the number of smart trash cans on the platform is nearly negligible compared to that of general trash cans. Thus the category refinement should significantly improve the matching outcomes of smart trash cans but have insignificant effects on general trash cans. With one heavily treated group (smart trash can) and one slightly treated group (general trash can), we can control the platform's other potential marketing efforts that may influence the search and matching outcomes.

We use consumers whose search queries are related to "smart trash can" as the treatment group. Our control groups come from several unaffected product categories that are close in the product attributes and consumer characteristics with the refined category. We use consumers whose search queries are related to "robot vacuum" as the control group in our main analysis and the others as robustness checks.[22] We compare the changes in consumers' matching outcomes and search intensity in the treatment group before and after the category refinement relative to the same changes over time among consumers in the control group.[23]

4.2 Summary Statistics

The shopping experience on e-commerce platforms can be summarized as search, click, and purchase. Figure 4 provides an illustration of these three stages. It starts with a consumer describes what she wants by putting

[20] Figure A1 in the appendix shows the daily number of consumers who search for smart trash cans on the platform. As we can see from the figure, the demand for smart trash cans is quite stable throughout our sample period.
[21] For example, seasonal fruits like watermelon demonstrate stronger demand during the summer.
[22] See section 6.4 for detail. We use category "air purifier", "vacuum" and "mop" as the alternative control group respectively.
[23] In Table A2 in the appendix, we compare consumers who search for smart trash cans before and after the category refinement. We find no significant differences in consumer characteristics such as purchase power or mobile app engagement across consumers in the smart trash can category before and after the category refinement.

in a query in the search box. Then the search engine interprets the query and returns the relevant products to the consumer. Next the consumer looks through the list of products in the search results and clicks on some of them which match her preferences to get further information. After comparing pros and cons of every item she has clicked, she may end up with buying one that she likes most. Our data records what consumers see and every decision they make throughout this journey.

[Figure 4 about here]

Our data for the analysis come directly from Alibaba and include around 7 million consumers over two-year period ending in December 2019. In our main analysis, we use search and purchase data four weeks before and after the refinement took effect.[24] To sample all searches that belong to a particular category, we first rank the top 100 popular search queries that consumers use within the category. Then we pull out all search records that are derived from these search queries. We only keep all search results derived from a user's first search query in the first search session that is related to the product category. The idea is to capture the very beginning of a search process and examine how the search precision in the initial search impacts consumers' exploration efforts in the later search refinement process.

The unit observation is at search level. Each search is defined by a unique user ID, a search query, and a specific date.[25] At the search level, we observe how many listings a consumer view and how many pages she scrolls down. We also capture a consumer's engagement with the platform by calculating the number of clicks she makes and total amount of time she spends on all clicked listings. Based on whether a consumer makes any clicks or any purchases after a search, we further generate two dummies to summarize the matching outcomes of search results. At the consumer level, we can observe each consumer's entire search and purchase histories throughout our sample period.[26] We know whether a consumer logs into the platform and has any engagement with it. Thus we can track consumers' responses several weeks after the changes in search algorithms. At the seller level, we observe a detailed transaction history of every item a seller has ever published. Therefore, we

[24]We lengthen the time window to be eight weeks before and after the category refinement in robustness checks.
[25] Each search session is identified by a unique session id. In the data, a starting time stamp and an ending time stamp can characterize a search session.
[26] All consumers are anonymous in our data. We do not use any personal information of consumers.

can directly calculate the transaction revenues a seller earns from each item. We also know the characteristics of each seller such as ratings and number of days since opening the store. These data allow us to link sellers' attributes with their strategic behaviors in response to the changes in search algorithms.

Table 2 reports summary statistics for matching outcomes and search intensity of consumers searching for the refined category "Smart Trash Can" and for the unaffected category "Robot Vacuum" before and after the category refinement. The average click through rate and purchase rate of both categories are higher in the after period. These differences, particularly the increase in the category "Smart Trash Can", could be a consequence of the platform category refinement. In addition, consumers who search for smart trash cans on average view and click fewer listings after the category refinement, while there is no significant change in search intensity for consumers who search for vacuum robots.

[Table 2 about here]

5. Empirical Strategy

5.1 Difference in differences

We provide several quasi-experimental research designs beginning with a difference-in-differences analysis of the impact of the category refinement. Specifically, we begin by estimating the following model.

$$y_{iqt} = \beta_0 + \beta_1 After_t \times SmartTrashCan_q + \gamma' X_{iqt} + \mu_q + \nu_t + \varepsilon_{iqt} \qquad (1)$$

Where y_{iqt} is a measure of search decisions or purchase decisions for consumer i who searches for query q on day t. Each observation in our sample is uniquely defined by a consumer ID, a search query, and a search date. We use search data from 3/18/2019 to 5/18/2019, and *After* equals to one for April 18, 2019 through May 18, 2019. X_{iqt} is a vector of covariates capturing consumer i's previous search and purchase behaviors and characterizing what consumer i see in the search results for query q on day t. For example, we create a dummy variable to indicate whether consumer i searches or purchases any related product categories such as mops and trash bags before the day t. We also calculate the average price of the search listings that consumer i sees when she searches for the query q on day t. Besides, we include search query fixed effects μ_q and week by month fixed effects ν_t.

The estimates for this and all subsequent models are weighted using the number of consumers at the query-by-week level. Drawing on Bertrand et al. (2004), we cluster the standard error at the query level to allow for correlation of errors over time within each of 200 search queries in our sample. We have also explored alternative levels of clustering, including: category level and query-by-month level. Statistical inference results are robust to these alternative clustering choices. In our first specification, we compare consumers whose search queries are related to the refined product category "Smart Trash Can" to those whose search queries are related to the unrefined product category "Robot Vacuum". β_1 measures the impact of the category refinement on consumers' search and purchase behaviors.

Identification of the difference-in-differences model requires that in the absence of the category refinement, the control group (consumers searching for robot vacuums) should have similar trends to the treated group (consumers searching for smart trash can). To explore the validity of the design, we do an "event time" analysis. This allows an examination of the pre-trends. We replace *After* × *Smart Trash Can* with a full set of week dummies interacted with *Smart Trash Can* in equation (2).

$$y_{iqt} = \beta_0 + \sum_{\delta \neq -1} \beta_\delta Time\_to\_Treat_\delta \times SmartTrashCan_q + \gamma' X_{iqt} + \mu_q + \nu_t + \varepsilon_{iqt} \quad (2)$$

5.2 Triple Differences

To rule out the alternative explanations such as potential selection bias in refining categories, we further include consumers whose search queries are related to "General Trash Can" as another treatment group and use a triple differences analysis (Goldfarb and Tucker 2011, Rishika et al. 2013) as a placebo test. The first part of the triple differences analysis is the difference in matching outcomes and search intensity between the two treatment groups (Smart Trash Can and General Trash Can) and the control group (Robot Vacuum) after the category refinement, minus the difference in matching outcomes and search intensity between the treatment groups and the control group before the category refinement. This is captured by β_1 in equation (3), and is a difference-in-differences analysis of the effects of the category refinement. The second part of the triple differences analysis, shown by $After_t \times TrashCan_q \times SmartTrashCan_q$ is the difference in matching outcomes and search intensity between two treatment groups, consumers searching for the niche category "Smart Trash Can" and ones searching for the broad category “Trash Can”, after the category refinement, minus the difference between in matching outcomes and search intensity in these two treatment groups before the category refinement.

$$y_{iqt} = \beta_0 + \beta_1 After_t \times TrashCan_q + \beta_2 After_t \times TrashCan_q \times SmartTrashCan_q + \gamma' X_{iqt} + \mu_q + \nu_t + \varepsilon_{iqt} \quad (3)$$

A difference-in-differences analysis controls for any omitted factors that influence consumers’ search and purchase decisions differently for the affected and the unaffected categories. The benefit of the triple differences analysis is that, in addition to control for those factors, it will remove confounders that are accompanied with the category “Trash Can” being the targeted category. For example, if Alibaba gives special attention to the category “Trash Can” and allocates more marketing expenses into it, we should see significant changes in both the subcategory “General Trash Can” and the subcategory “Smart Trash Can” after the refinement. However, if becoming the targeted category per se does not affect consumer behaviors, the category refinement should significantly improve the matching outcomes of consumers who search for smart trash cans but have insignificant effects on consumers who search for general trash cans.

6. Results

6.1 Matching Outcomes

We begin by presenting results for the difference-in-differences model using the refinement of the category "Trash Can". After the refinement, the product category "Trash Can" was divided into two separate categories: "General Trash Can" and "Smart Trash Can." Our treatment group for the main estimates is the consumers who search for smart trash cans during our sample period, while the control group is those who search for robot vacuums during the same period. We use two dummy variables to capture consumers' responses to what they see after they put in a search query. The first one indicates whether a consumer makes at least one click after being presented with the search listings. The other one indicates whether a consumer makes at least one purchase during the first search session. These two variables generally summarize whether a consumer is interested in or satisfied with the search results that the search engine delivers according to her search keyword.[27] We expect that consumers are more likely to make a click and even end up with a purchase when the search results well match their preferences.

Results from the difference-in-differences model are shown in the first two columns of Table 3. Each column represents the estimates from a separate regression with a different dependent variable. The first column indicates that consumers who search for smart trash cans, relative to those who search for robot vacuums, are 0.69 percentage points more likely to make a click after the category refinement, translating into 37.3% increase in the average click-through rate. The purchase rate in the treatment group also increases by roughly 36.1% after the category refinement, resulting in 64.4% increase of gross merchandise volume in the category "Smart Trash Can."[28]

[Table 3 about here]

As a placebo test, we estimate the triple differences model from equation (3). We expand our sample by adding consumers who search for general trash cans as another treatment group into our existing sample. The results are reassuring. As shown in the last two columns of Table 2, the coefficients of $After_t \times TrashCan_q$ are not statistically significant but the coefficients of $After_t \times TrashCan_q \times SmartTrashCan_q$ are positive and significant. The results suggest that for consumers who search for general trash cans, neither the click-through rate or the purchase rate significantly changes after the category refinement. Thus the category refinement has negligible effect on these consumers, alleviating the concern of the selection bias in the category refinement.

To explore the validity of the difference-in-differences design, we use an "event time" analysis. This allows an examination of the pre-trends. In Figure 5, we plot the week by $SmartTrashCan_q$ interactions using estimation from equation (2), where we leave out week 16 as the reference point. Prior to the category refinement, we find little evidence of differential group trends. For week dummy smaller than 16, most treatment coefficients are less than 0.005 points in magnitude and seldom reach statistical significance. After

[27] Click-through rates directly reflect whether consumers are interested in the product listings in the search results. While purchase-rates also depend on the quality of search results, they are affected by other factors such as product prices.

[28] The purchase rate refers to the percentage of consumers who purchase at least one smart trash can during a search session. We don't count purchases in other product categories. Similarly, in the control group, we only calculate the purchases of robot vacuums.

the category refinement, click-through rates and purchase rates increase significantly among consumers who search for smart trash can.[29]

[Figure 5 about here]

6.2 Search Intensity

Besides matching outcomes, we further explore how the improvement in search quality affects consumer search intensity. We use the same difference-in-differences model with another set of dependent variables to examine how consumers' search intensity changes due to the category refinement. We develop three measures to describe consumers' search intensity during a search session. The first is how many listings a consumer views after putting in a search query. The key decision a consumer makes in this viewing process is whether to scroll down or not. If she is particularly interested in one of the listings in the search results, she needs to obtain additional information by clicking on it. In light of this, the other two measures focus on how many listings a consumer clicks during the viewing process and the total time she spends on these clicked listings. These two measures represent the search on the extensive margin and the search on the intensive margin respectively (Ursu et al., 2019).

Similar to Table 3, Table 4 presents results from for the difference-in-differences model and the triple differences model. We restrict our sample to consumers who make at least one click during a search session.[30] The first column of Table 4 shows results with the dependent variable being number of listings viewed by a consumer. The other two columns estimate the same model using the number of clicked listings and the total time a consumer spends on these clicked listings as the dependent variable respectively. We take logs for all these three dependent variables when running regressions.

[29] In Figure 5, the click-through rate and the purchase rate show a slightly decrease after an initial jump. This can be explained by the long-term changes from the supply side. In section 7.3, we find that small and new sellers are more likely to enter the narrow category "Smart Trash Can" after the category refinement. Consumers may be hesitant to click and purchase products from sellers with fewer ratings (Ghose et al 2014, Dinerstein et al 2018).

[30] This restriction can help us generate meaningful variations in consumer search intensity. As we can see from the summary statistics in Table 2, more than 95% consumers do not make any clicks during a search session. By restricting our sample to consumers with positive clicks, we can examine whether consumers decrease search intensity when they find what they want in the search results.

Results from Table 4 indicate that consumers reduce search intensity after the category refinement. As we can see from the estimates of the difference-in-differences model, for consumers who clicks on listings, the total number of viewed listings decreases by 4.4% after the category refinement. The category refinement also makes consumers click 3.7% fewer listings and spend 6.6% less time on clicking. Our results from the triple differences model also suggest that the effect of category refinement on search intensity is mainly significant for consumers who search for smart trash cans.[31] Our findings are consistent with the theoretical prediction in Yang (2013). His model predicts that the overall search of consumers could possibly decrease if there is an increase in search quality. As search precision improves, consumers are more likely to meet the right product and end up purchasing more quickly during the search process.

[Table 4 about here]

6.3 Consumer Heterogeneity

To understand how matching outcomes and search intensity vary across different consumers, we have estimated the heterogeneous treatment effects of the category refinement. In Table 3 and Table 4, we group all consumers who search for a product category together and get a general estimate of the average treatment effect for all consumers. Since each category consists of several related search queries, in this section, we further zoom into the specific search query during each search session to understand how consumers' shopping intents behind these search queries moderate the impact of the category refinement.

We classify all search queries that are related to the category "Smart Trash Can" into two subgroups: general interest and specific need. Table A3 in the appendix provides examples of these two groups of search queries. The "general interest" subgroup only includes search queries with the name of a product category or an imprecise description of desirable features of a product (Yang et al 2014). For example, "smart trash can"

[31] Column 4 of Table 4 indicates that consumers who search for general trash cans are likely to view fewer listings after the category refinement. One possible explanation is that these consumers may stop searching earlier as they see more general trash cans in the search results. But the last two columns in Table 4 suggest that the decrease in search intensity for consumers searching for general trash cans is not significant when we consider two alternative measures regarding clicking. Clicking induces higher search costs and demands longer engagement than viewing (Chen and Yao 2017, Ursu et al 2019).

and "sensor motion trash can" are the two most searched queries in the category "Smart Trash Can". 30.49% consumers in our sample use these two generic queries when searching for smart trash cans. These consumers only have a category of products in mind and are often at the early stage of shopping. They learn and adjust their preferences as exploring more products in the search results. We identify the "specific need" subgroup as consumers whose search queries which either contain specific brand names or express particular application scenarios. One example is the query "Smart Trash Can+Automatic Packaging+Perfect for Home." Consumers should know exactly which products to buy at the time when they put this query into the search box. They have gathered enough information about smart trash cans and are ready to land a deal. After grouping all search queries into two subgroups, we replace *After* × *Smart Trash Can* with two dummies interacted with *After* × *Smart Trash Can*. These two dummies indicate which subgroup a search query belongs to. We estimate how treatment effects vary across different subgroups of queries.

Table 5 presents results from this exercise. As for the matching outcomes, the first two columns indicate that the category refinement significantly increases click through rates for two subgroups of consumers. This positive matching effect is especially pronounced for consumers with specific shopping needs. After the category refinement, this subgroup of consumers are 3.2% more likely to make at least one click after viewing the search results.[32] Interestingly, as suggested by the next three columns, consumers with general shopping interests show significantly lower search intensity as a result of the category refinement. This subgroup of consumers view 4.3% fewer listings, make 3.7% fewer clicks, and spend 6.2% less time on clicking. As for consumers with specific shopping needs, however, we don't find significant changes in their search intensity after the category refinement.

Overall, our results illustrate a tradeoff between exploitation and exploration in e-commerce search engine design. As search quality improves, at the extensive margin, more consumers will easily find products that match their preferences and then make a purchase. On the other hand, at the intensive margin, the more refined search results will discourage consumers' self-exploration efforts and thus make them view and click

[32]The purchase rates for consumers with specific shopping needs do not significantly increase after the category refinement. This is partly because it takes longer time for consumers to finalize a purchase. Since our unit observation is at search session level, we consider our estimates as a lower bound for purchase rates.

fewer products before landing a deal. The positive matching effect is more pronounced for consumers with specific shopping intents, while the negative search deterrence effect is disproportionately on consumers with general shopping interests.

[Table 5 about here]

6.4 Robustness Checks

We conduct extensive robustness checks for the above results in Table A4 and A5 and our results are robust to alternative control groups and alternative samples.

More specifically, to alleviate the concern of comparability of the control group, we re-estimate our difference-in-differences model using the category "Air Purifier", "Vacuum" and "Mop" as the alternative control group respectively while maintaining the category "Smart Trash Can" as the treatment group. These unaffected product categories are candidates for our control groups for three reasons. First, the characteristics of consumers searching for these product categories are close to those of consumers in our treatment group. We identify these product categories by looking into the search records of consumers who search for smart trash cans. "Air Purifier", "Vacuum" and "Mop" are most frequently searched categories by these consumers. Second, they all share the similar application scenarios with smart trash cans. They are all used in the home for cleaning purposes and are complements with each other. Third, search and purchase outcomes in these product categories have similar trends with the treated group during our pre-treatment period, thus meeting the identifying consumption of the difference-in-differences model. We get similar estimates of the effects of the category refinement on matching outcomes and search intensity as we see in Table 3 and Table 4.

We also re-estimate our model using alternative samples to address the following concerns. First, our findings may be merely a manifest of seasonal promotions if other marketing efforts are accompanied with the category refinement during our sample period, i.e., 3/18/2019-5/18/2019. Thus we estimate the model using the sample from last year, i.e., 3/18/2018-5/18/2018. We find that there is no significant change in either click through rates or purchase rates for consumers who search for smart trash cans. Hence our findings are mainly driven by the changes in the search algorithms. In addition, someone may worry that the searchers of smart trash cans are not comparable to those in the control groups we mention above. As a response to this concern, we restrict our attention to the consumers who search both "smart trash cans" and "robot vacuums." This leaves us a "common-user" subsample of around 0.18 million consumers. The idea is to control unobserved consumer characteristics in the treatment group that may result in differences in search and purchase behaviors compared with the control group. We find that controlling for this factor in our analysis has no impact in our results. Lastly, we estimate a sample of four-month time window, i.e., 2/18/2019 – 6/18/2019. We find that our results still remain when we lengthen the pre-treatment window in our difference-in-differences model.

7. The Long-Run Effects of Improving Search Precision

In the main analysis, we have documented that increasing search precision helps consumers find matched products more quickly but substantially reduces their exposure to other products on the platform. One follow-up question is what are the long-run effects of improving search precision. In this section, we examine the impacts of the category refinement on consumer engagement and unplanned purchases over a longer period of time. Previous studies on in-store shopping behaviors have suggested that increasing travel distances within a store can increase consumer unplanned spending by exposing them to more products (Silley et al 2010, Hui et al 2013). [33]These insights can be applied in the online settings, where imprecise search results may make consumers stay longer and explore more on the platform, resulting in more spending.

7.1 Consumer Engagement

We track consumers' activities on the platform for one week after their initial search session for smart trash cans (treatment group) or robot vacuums (control group). Our long search panel data enable us to examine how consumers' engagement with the platform change after seeing the more precise search results. We use the number of search sessions and the number of search queries to describe consumer search activities in the following week after the initial search session. Besides these two metrics, we count the number of days each

[33] Offline retailers have come up with marketing strategies that deliberately increase consumers' search frictions, such as scattering popular product categories throughout the store (Granbois 1968, Iyer 1989).

consumer logs into the Taobao app during the same time period. We further classify consumers into two groups based on their average time spent in a search session before category refinement. Table A6 in the appendix summarizes the average length of consumer search sessions in our sample. We define consumers whose clicking time is above the median as *exploratory searchers* who tend to scan and browse in a search environment without predefined goals. The rest of consumers are *goal-directed searchers* who are guided by specific goals and are motivated to gather information efficiently. Based on these two types of consumers, we study the heterogeneous effects of the category refinement on consumer search and purchase behaviors in the long run.

Table 6 presents the difference-in-differences model estimation similar to our main analysis with the dependent variables being as the three measures of consumer engagement with the platform. Results indicate that improving search precision can increase the satisfaction of goal-directed searchers at the cost of losing the attention of exploratory searchers. After the category refinement, goal-directed searchers in the treatment group conduct more searches, use more search queries, and log into the platform more times in the following week after seeing the more precise search results. The positive experiences of quickly getting what they want make them more likely to return to the platform for shopping. In contrast, we find a significant decrease in engagement with the platform for exploratory searchers when the search precision increases. Their average number of days visiting the platform during a week decreases by 4.1%, their number of searches in a week decreases by 5.5%, and their number of search queries decreases by 5.6%. These consumers mostly enjoy browsing and scanning various product categories on the platform. For this type of consumers, too precise search results may take away the fun of exploration and leave them spend less time on the platform.

[Table 6 about here]

7.2 Cross Selling

The precision of search results in one category can impact consumers' spending in other related categories. To measure the impact of the category refinement on cross-selling opportunities, we calculate the total purchase amount across other categories in the week following the initial search. For consumers in the treatment group, the relevant product categories include mops, trash bags, vacuums, air purifiers, and storage racks. These are among the mostly searched categories based on treated consumers' search histories. We also calculate the total purchase amount across the leaf categories that share the same first-tier or second-tier category with the category "Smart Trash Can." [34] We use these two groups of categories as alternative measures of the relevant categories that could be affected by the precision of search results in the treatment group. Similarly, we calculate the total amount of purchases in the relevant categories for consumers in the control group. We estimate the same model with another set of dependent variables depicting consumer spending.

As suggested in Table 7, we find significant changes in consumers' purchases of the product categories that are related to the category "Smart Trash Can" after the category refinement. Improving search precision can boost goal-directed searchers' confidence on searching on the platform, thus increasing their spending in other product categories in the long run. However, e-commerce platforms may fail to make the most sales of exploratory searchers if the search results are too precise. Results in Table 7 indicate that these consumers are likely to decrease their spending in the relevant products by about 1-2% in the following week when the search precision improves. We regard our estimates as a lower bound of the long-run effect of the category refinement since it may take several weeks for some consumers to finalize a purchase. Imprecise search results bring serendipity, but the exact time that this serendipity leads to purchases can vary across consumers. This phenomenon is especially true in the environment of online shopping as consumers do not face the same time constraint as they do in a physical mall or a shopping center. [35]

[Table 7 about here]

8. Discussion

In this section, we first discuss the underlying mechanism that explains the changes in matching outcomes and search intensity after the category refinement. Then, we document third party sellers' strategic entry behaviors in response to the category refinement.

[34] "Smart Trash Can" is on the category note "Home/Personal Cleaning Tools -> Home / Floor Cleaning Tools -> Smart Trash Can". Categories that share the second-tier category "Home / Floor Cleaning Tools" with "Smart Trash Can" include barrels, basins, brooms, and cleaning cloths. Categories that share the first-tier category "Home/Personal Cleaning Tools" include combs, toothbrushes, and shavers.

[35] Previous studies on in-store shopping document that longer travel distances make consumers increase unplanned purchases (Silley et al 2010, Hui et al 2013). When consumers shop online, their search activities can span several weeks, with most search sessions ending in no purchase (Blake et al 2016).

7.1 Mechanism: Targeted Allocation of Search Traffic

Search engines help consumers find products by indexing products for relevant keywords. If a product is indexed for a keyword, it will show as a result when a customer uses that keyword in the search bar. The primary method of getting indexed for a keyword on e-commerce platforms is by selecting a relevant category. If a category is indexed with a search query, all products in the category as a whole will receive relatively more search traffic from that search query.

To test this hypothesis, we calculate the search traffic of the query "smart trash can" distributed to the category "Smart Trash Can" and the category "Trash Can" before and after the category refinement. We define the search traffic of a category as the total count of how many times consumers view the products belonging to that category.[36] Figure 6 shows the ratios of weekly search traffic that is distributed to these two categories respectively for the query "Smart Trash Can" throughout 2019. Before the category refinement (April 18, 2019), less than 2% of search traffic was allocated to the category "Smart Trash Can". [37]After the category refinement, this number rose to over 15% and was increasing to nearly 25% at the end of 2019.

These results suggest that the search engine gives more targeted search traffic to the refined product category. Before the category refinement, consumers who search for a niche product like smart trash cans, are more likely to see only general trash cans in the search results. After the change, with the "smart trash can" being singled out as an independent category, the search engine would preferentially deliver products from the category "Smart Trash Can" to consumers who search for smart trash cans. With more refined and targeted search results by a boosting search engine, consumers have higher chances of meeting the right products.

[Figure 6 about here]

7.2 Supply Side Response: Stronger Long Tail Effect

The previous analysis mainly focuses on the reactions from the demand side, trying to understand how the improvement in search quality affects consumers' matching outcomes, search intensity, and engagement with

---

[36] This measure is also called page views(PV) in internet advertising (Danaher 2007). An alternative measure of search traffic is the number of unique visitors(UV) that have viewed a product in a given period of time (Gallion and Moreno 2014, Sun et al 2020). We get similar results when using these two measures of search traffic.

[37] Alibaba created the subcategory "Smart Trash Can" on March 27, 2019 and advertised for this new option. It took the platform around three weeks to change the search algorithm and make the *de facto* category refinement on April 18, 2019.

the platform. But sellers could also re-optimize their competition strategies in the long run as a result of the changes in search algorithms. Thus in this part, we look at the supply side of the market and examine whether sellers have strategic behaviors after the category refinement.

We find that the weekly number of smart trash cans on the platform increases by 150% as search quality improves. The more targeted search traffic attracts many more sellers to put their products into this niche category. We further pull out all sellers who show up in the search results of the query "smart trash can" and divide them into two groups: sellers who enter the niche category "Smart Trash Can" and sellers whose products belong to the category "General Trash Can". Figure 7 suggests that sellers with fewer ratings are more likely to list products in the "Smart Trash Can" category. Small and new sellers with lower operating costs can earn a reasonable profit if entering a narrow category. In contrast, large sellers often hire a team of specialists to help operate their e-commerce stores. Entering a narrow category with limited search traffic may not generate enough profits to cover their monthly salary expenses.

[Figure 7 about here]

Niche product sellers as a whole will gain more profits as the quality of search improves. We document that both total transaction volume and revenues of smart trash cans significantly increase after the category refinement. By reducing the search friction on the platform, the search engine makes it easier for relevant sellers to reach consumers who may have higher willingness to pay for their products. As a result, the total revenues of the refined product category nearly doubled. Sellers who sell niche products are significantly better off with the advance of the search technology. With the increasing number of specialized products that cater to more specific segments of consumers, small and new sellers gain more market shares and the distribution of sales will become flatter.[38]

[38] Our findings are consistent with the literature on how online search tools or recommendation systems affect sales distribution (Brynjolfsson et al 2011, Bar-Isaac et al.2012, Hervas-Drane 2015, Chen and Yao 2017). For example, Hervas-Drane(2015) builds a search model to show that a personalized recommendation system tends to reduce the concentration of sales. The recommendation system increases the matching between consumers and products and thus make consumer search more targeted. The same rationale can be applied to our setting.

9. Conclusion and Implications

This paper explores the role of the e-commerce platform in improving search precision and measure the value of doing so to consumers with different shopping needs. In collaboration with Alibaba, we exploit a particular change in search design for our identification: category refinement. With a difference-in-differences analysis combined with an event study estimation and a triple differences analysis, we find that the matching between consumers and relevant products significantly increases as search precision improves. This positive matching effect is more significant for consumers who know exactly which products to buy. However, the refined search results decrease consumers' search activities and their engagement with the platform in the long run. This negative search deterrence effect is more pronounced for exploratory searchers who only have a category of products in mind and are at the early stage of online shopping. Overall, our results imply a trade-off between exploitation and exploration in the e-commerce search design.

Our findings provide important implications for e-commerce platforms in general. With cutting-edge machine learning techniques and the availability of big data, researchers and engineers in computer science have been working on the designs of product rankings and category refinements for the improvement in information retrieval system results (Dumais and Chen 2000, Xing et al 2008, Bilal 2012, Farhoodi et al 2013). Most of these state-of-art algorithms can immediately boost search engine revenues but may be unsustainable in a longer time horizon (Anderson et al 2020). From the perspective of economics, our study suggests that platforms should augment accuracy with diversity, novelty, and serendipity when considering the metrics of search algorithms. By encouraging consumers to explore diverse, novel, and serendipitous product categories, e-commerce platforms can increase consumer engagement and generate long-term impacts of user value. In addition, our results indicate that e-commerce companies should fully leverage the textual data from online search queries to understand consumers' shopping intents. Product search engines are quite different from information search engines such as Google or Bing (Ghose et al 2014). On e-commerce platforms, people don't ask questions. Instead, they search for products. The semantic match of search queries is not enough. Platforms should extrapolate what products consumers really want when they type in these search queries.

Our study of platform search design has limitations that imply several directions of future research. First, while we have illustrated the exploitation-exploration tradeoff, the optimal design of search precision and the optimal depth of categorization scheme on e-commerce platforms are both beyond the scope of this paper. Our results indicate that search algorithms impact firms' strategic decisions such as category entry. Future work in this direction is required to take firms' incentives into account when designing search algorithms. Second, our analysis provides reduced-form evidences of the impacts of improving search precision on consumer exploratory search behaviors, abstract away from modeling the sequential path of the search refinement process. It is interesting to understand how consumers refine searching objects from initial broad categories all the way to final targeted products and how platforms use search or recommendation algorithms to guide this discovery process. Finally, though as complete and granular as possible, our search data are confined to one particular domain, failing to capturing search activities across other e-commerce websites as well as offline retailers. Data sets that track individual consumers' entire search path both online and offline will further enrich our insights and will enable additional discussions such as how competition from other platforms or consumer multi-homing moderates the impacts of platform search design.

References

Adamopoulos, P., and Tuzhilin, A., 2014. On unexpectedness in recommender systems: or how to better expect the unexpected. *ACM Transactions on Intelligent Systems and Technology* 5(4): 54:1-32.

Aghion, P., Bolton, P., Harris, C. and Jullien, B., 1991. Optimal learning by experimentation. *The Review of Economic Studies*, *58*(4), pp.621-654.

Aggarwal, C.C., 2018. *Machine learning for text*. Cham: Springer International Publishing.

Alexandrov, A., and Koulayev, S. 2017. No shopping in the US mortgage market: Direct and strategic effects of providing information. *Working paper*

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

Anderson, A., Maystre, L., Anderson, I., Mehrotra, R. and Lalmas, M., 2020, April. Algorithmic effects on the diversity of consumption on spotify. In *Proceedings of The Web Conference 2020* (pp. 2155-2165).

Bakos, Y., 2001. The emerging landscape for retail e-commerce. *Journal of Economic Perspectives*, *15*(1), pp.69-80.

Basker, E., Klimek, S. and Hoang Van, P., 2012. Supersize it: The growth of retail chains and the rise of the "Big-Box" store. *Journal of Economics & Management Strategy*, *21*(3), pp.541-582.

Bar-Isaac, H., Caruana, G. and Cuñat, V., 2012. Search, design, and market structure. *American Economic Review*, *102*(2), pp.1140-60.

Baye, M.R. and Morgan, J., 2001. Information gatekeepers on the internet and the competitiveness of homogeneous product markets. *American Economic Review*, *91*(3), pp.454-474.

Benner, M.J. and Tushman, M.L., 2003. Exploitation, exploration, and process management: The productivity dilemma revisited. *Academy of Management Review*, *28*(2), pp.238-256.

Bertrand, M., Duflo, E. and Mullainathan, S., 2004. How much should we trust differences-in-differences estimates? *The Quarterly Journal of Economics*, *119*(1), pp.249-275.

Bilal, D., 2012. Ranking, relevance judgment, and precision of information retrieval on children's queries: Evaluation of Google, Yahoo!, Bing, Yahoo! Kids, and ask Kids. *Journal of the American Society for Information Science and Technology*, *63*(9), pp.1879-1896.

Blake, T., Nosko, C. and Tadelis, S., 2016. Returns to consumer search: Evidence from eBay. In *Proceedings of the 2016 ACM Conference on Economics and Computation* (pp. 531-545).

Burdett, K. and Judd, K.L., 1983. Equilibrium price dispersion. *Econometrica*, pp.955-969.

Bubeck, S. and N. Cesa-Bianchi., 2012. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends in Machine Learning* 5 (1), 1– 122

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), pp.693-712.

Brynjolfsson, E. and Smith, M.D., 2000. Frictionless commerce? A comparison of Internet and conventional retailers. *Management science*, *46*(4), pp.563-585.

Brynjolfsson, E., Hu, Y. and Simester, D., 2011. Goodbye pareto principle, hello long tail: The effect of search costs on the concentration of product sales. *Management Science*, *57*(8), pp.1373-1386.

Chen, M.K., 2016, July. Dynamic pricing in a labor market: Surge pricing and flexible work on the uber platform. In *Proceedings of the 2016 ACM Conference on Economics and Computation* (pp. 455-455).

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

Chiou, J.S. and Ting, C.C., 2011. Will you spend more money and time on internet shopping when the product and situation are right? *Computers in Human Behavior*, *27*(1), pp.203-208.

Currie, J.M. and MacLeod, W.B., 2020. Understanding doctor decision making: The case of depression treatment. *Econometrica*, *88*(3), pp.847-878.

Danaher, P.J., 2007. Modeling page views across multiple websites with an application to internet reach and frequency prediction. *Marketing Science*, *26*(3), pp.422-437.

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), pp.2955-80.

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

Dinerstein, M., Einav, L., Levin, J. and Sundaresan, N., 2018. Consumer price search and platform design in internet commerce. *American Economic Review*, *108*(7), pp.1820-59.

Dong, X., Morozov, I., Seiler, S. and Hou, L., 2019. Estimation of Preference Heterogeneity in Markets with Costly Search. *Working paper*.

Dumais, S. and Chen, H., 2000, July. Hierarchical classification of web content. In *Proceedings of the 23rd annual international ACM SIGIR conference on Research and development in information retrieval* (pp. 256-263).

Dzyabura, D. and Hauser, J.R., 2019. Recommending products when consumers learn their preference weights. *Marketing Science*, *38*(3), pp.417-441.

Einav, L., Farronato, C. and Levin, J., 2016. Peer-to-peer markets. *Annual Review of Economics*, *8*, pp.615-635.

Eliaz, K. and Spiegler, R., 2016. Search design and broad matching. *American Economic Review*, *106*(3), pp.563-86.

Fleder, D. and Hosanagar, K., 2009 Blockbuster culture's next rise or fall: the impact of recommender systems on sales diversity. *Management Science* ,55(5):697-712.

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

Farhoodi, M., Shiry Ghidary, S. and Yari, A., 2013. Quality Improvement of Information Retrieval System Results by Document Classification. *International Journal of Information and Communication Technology Research*, *5*(3), pp.45-54.

Fradkin, A., 2017. Search, matching, and the role of digital marketplace design in enabling trade: Evidence from Airbnb. *Working Paper*.

Gallino, S. and Moreno, A., 2014. Integration of online and offline channels in retail: The impact of sharing reliable inventory availability information. *Management Science*, *60*(6), pp.1434-1451.

Galenianos, M., & Gavazza, A. (2017). A structural model of the retail market for illicit drugs. *American Economic Review*, *107*(3), 858-96.

Gardete, P., and Hunter A,M. (2019) Guiding consumers through lemons and peaches: A dynamic model of search over multiple characteristics. *Working paper*

Ge M, Delgado-Battenfeld C, Jannach D (2010) Beyond accuracy: Evaluating recommender systems by coverage and serendipity. *RecSys '10. Proceedings of the 2008 ACM Conference on Recommender Systems*, Barcelona, Spain. 257–260

Ghose, A., Ipeirotis, P.G. and Li, B., 2012. Designing ranking systems for hotels on travel search engines by mining user-generated and crowdsourced content. *Marketing Science*, *31*(3), pp.493-520.

Ghose, A., Ipeirotis, P.G. and Li, B., 2014. Examining the impact of ranking on consumer behavior and search engine revenue. *Management Science*, *60*(7), pp.1632-1654.

Ghose, A., Ipeirotis, P.G. and Li, B., 2019. Modeling consumer footprints on search engines: An interplay with social media. *Management Science*, *65*(3), pp.1363-1385.

Goldfarb, A. and Tucker, C.E., 2011. Privacy regulation and online advertising. *Management Science*, *57*(1), pp.57-71.

Granbois, D.H., 1968. Improving the study of customer in-store behavior. *Journal of Marketing*, *32*(4), pp.28-33.

Gu, N. and Wang Y.K., 2019. Consumer online search with partially revealed information. *Working paper*

Groysberg, B. and Lee, L.E., 2009. Hiring stars and their colleagues: exploration and exploitation in professional service firms. *Organization Science*, *20*(4), pp.740-758.

Hagiu, A. and Wright, J., 2020. Platforms and the exploration of new products. *Management Science*, *66*(4), pp.1527-1543.

Hervas-Drane, A., 2015. Recommended for you: The effect of word of mouth on sales concentration. *International Journal of Research in Marketing*, *32*(2), pp.207-218.

Hitsch, G.J., Hortaçsu, A. and Ariely, D., 2010. Matching and sorting in online dating. *American Economic Review*, *100*(1), pp.130-63.

Hui, S.K., Inman, J.J., Huang, Y. and Suher, J., 2013. The effect of in-store travel distance on unplanned spending: Applications to mobile promotion strategies. *Journal of Marketing*, *77*(2), pp.1-16.

Holland, J.H., 1975. Adaptation in natural and artificial systems: an introductory analysis with applications to biology, control, and artificial intelligence. *University of Michigan Press.*

Horton, J.J., 2019. Buyer uncertainty about seller capacity: Causes, consequences, and a partial solution. *Management Science*, *65*(8), pp.3518-3540.

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

Honka, E. and Chintagunta, P., 2017. Simultaneous or sequential? search strategies in the us auto insurance industry. *Marketing Science*, *36*(1), pp.21-42.

Honka, E., Hortaçsu, A., & Wildenbeest, M. 2019. Empirical search and consideration sets. In *Handbook of the Economics of Marketing* (Vol. 1, pp. 193-257). North-Holland.

Hortaçsu, A. and Syverson, C., 2004. Product differentiation, search costs, and competition in the mutual fund industry: A case study of S&P 500 index funds. *The Quarterly Journal of Economics*, *119*(2), pp.403-456.

Hwang, M. and Park, S., 2016. The impact of walmart supercenter conversion on consumer shopping behavior. *Management Science*, *62*(3), pp.817-828.

Iyer, E.S., 1989. Unplanned purchasing: knowledge of shopping environment and. *Journal of Retailing*, *65*(1), p.40.

Janiszewski, C., 1998. The influence of display characteristics on visual exploratory search behavior. *Journal of Consumer Research*, *25*(3), pp.290-301.

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

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

Kuksov, D., 2004, "Buyer search costs and endogenous product design", *Marketing Science*, 23(4), 490–499

Kuran, T., 1988. The tenacious past: Theories of personal and collective conservatism. *Journal of Economic Behavior & Organization*, *10*(2), pp.143-171.

Lai, T.L. and Robbins, H., 1985. Asymptotically efficient adaptive allocation rules. *Advances in Applied Mathematics*, *6*(1), pp.4-22.

Larson, N., 2013, "Niche products, generic products and consumer search." *Economic Theory*, 52(2), 793-832

Li, B., Li, X. and Liu, H., 2017. Consumer preferences, cannibalization and competition: Evidence from the personal computer industry. *MIS Quarterly,* 42(2), pp.661-678

Liu, J. and Toubia, O., 2020. Search query formation by strategic consumers. *Quantitative Marketing and Economics*, *18*(2), pp.155-194.

Luco, F., 2019. Who benefits from information disclosure? The case of retail gasoline. *American Economic Journal: Microeconomics*, *11*(2), pp.277-305.

March, J.G., 1991. Exploration and exploitation in organizational learning. *Organization science*, *2*(1), pp.71-87.

McInerney, J., Lacker, B., Hansen, S., Higley, K., Bouchard, H., Gruson, A. and Mehrotra, R., 2018, September. Explore, exploit, and explain: personalizing explainable recommendations with bandits. In *Proceedings of the 12th ACM Conference on Recommender Systems* (pp. 31-39).

McCall, J.J., 1970. Economics of information and job search. *The Quarterly Journal of Economics*, pp.113-126.

McNee, S., Riedl, J., Konstan, J., 2006. Accurate is not always good: how accuracy metrics have hurt recommender systems. CHI EA 2006, *Extended Abstracts on ACM Human Factors in Computing Systems*, Quebec, Canada, 1097-1101.

Moraga-González, J. L., Sándor, Z., & Wildenbeest, M. R. 2018. Consumer search and prices in the automobile market. *Working paper*

Nishida, M., & Remer, M. 2018. The determinants and consequences of search cost heterogeneity: evidence from local gasoline markets. *Journal of Marketing Research*, 55(3), 305-320.

Pfeiffer, J., Pfeiffer, T., Meißner, M. and Weiß, E., 2020. Eye-tracking-based classification of information search behavior using machine learning: evidence from experiments in physical shops and virtual reality shopping environments. *Information Systems Research*.

Rishika, R., Kumar, A., Janakiraman, R. and Bezawada, R., 2013. The effect of customers' social media participation on customer visit frequency and profitability: an empirical investigation. *Information Systems Research*, *24*(1), pp.108-127.

Rhodes, A. and Zhou, J., 2019. Consumer search and retail market structure. *Management Science*, *65*(6), pp.2607-2623.

Rhodes, A., Watanabe, M. and Zhou, J., 2020. Multiproduct Intermediaries. *Journal of Political Economy*.

Schumpeter, J. A. 1934. The theory of economic development. *Harvard: Economic Studies*, *46*, pp.1600-0404.

Stahl, D.O., 1989. Oligopolistic pricing with sequential consumer search. *The American Economic Review*, pp.700-712.

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

Sun, H., Fan, M. and Tan, Y., 2020. An empirical analysis of seller advertising strategies in an online marketplace. *Information Systems Research*, *31*(1), pp.37-56.

Sutton, R.S. and Barto, A.G., 1998. Reinforcement learning: An introduction. *MIT press*.

Sorensen, A.T., 2000. Equilibrium price dispersion in retail markets for prescription drugs. *Journal of Political Economy*, *108*(4), pp.833-850.

Sørensen, J.B. and Stuart, T.E., 2000. Aging, obsolescence, and organizational innovation. *Administrative Science Quarterly*, *45*(1), pp.81-112.

Stilley, K.M., Inman, J.J. and Wakefield, K.L., 2010. Planning to make unplanned purchases? The role of in-store slack in budget deviation. *Journal of Consumer Research, 37*(2), pp.264-278.

Tam, K.Y. and Ho, S.Y., 2006. Understanding the impact of web personalization on user information processing and decision outcomes. *MIS quarterly*, pp.865-890.

Tan, X., Wang, Y. and Tan, Y., 2019. Impact of live chat on purchase in electronic markets: The moderating role of information cues. *Information Systems Research*, *30*(4), pp.1248-1271.

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

Ursu, R.M., Wang, Q. and Chintagunta, P.K., 2019. Search duration. *Working Paper*.

Varian, H.R., 1980. A model of sales. *The American economic review*, *70*(4), pp.651-659.

Vargas, S. and Castells, P., 2011. Rank and relevance in novelty and diversity metrics for recommender systems. In *Proceedings of the fifth ACM conference on Recommender systems* (pp. 109-116).

Weitzman, M.L., 1979. Optimal search for the best alternative. *Econometrica*, pp.641-654.

Wildenbeest, M. R. 2011. An empirical model of search with vertically differentiated products. *The RAND Journal of Economics*, *42*(4), 729-757.

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

Wolfinbarger, M. and Gilly, M.C., 2001. Shopping online for freedom, control, and fun. *California Management Review*, *43*(2), pp.34-55.

Xing, D., Xue, G.R., Yang, Q. and Yu, Y., 2008. Deep classifier: automatically categorizing search results into large-scale hierarchies. In *Proceedings of the 2008 International Conference on Web Search and Data Mining* (pp. 139-148).

Yang, H. 2013. Targeted search and the long tail effect. *The RAND Journal of Economics*, 44(4), 733-756.

Yang, S., Lu, S. and Lu, X., 2014. Modeling competition and its impact on paid-search advertising. *Marketing Science*, *33*(1), pp.134-153.

Yoganarasimhan, H., 2020. Search personalization using machine learning. *Management Science*, *66*(3), pp.1045-1070.

Zhang, S., Yao, L., Sun, A. and Tay, Y., 2019. Deep learning based recommender system: A survey and new perspectives. *ACM Computing Surveys (CSUR)*, *52*(1), pp.1-38.

Zhong, Z. 2019. A Model of Search Design. *Working Paper*

Zhou, T., Kuscsik, Z., Liu, J.G., Medo, M., Wakeling, J.R. and Zhang, Y.C., 2010. Solving the apparent diversity-accuracy dilemma of recommender systems. *Proceedings of the National Academy of Sciences*, *107*(10), pp.4511-4515.

Tables and Figures

Table 1: Comparison with Search Data in Recent Literature

	Chen and Yao(2017)	Dinerstein et al (2018)	Ghose et al (2019)	Dong et al (2019)	Gu and Wang (2019)	Ursu et al (2019)	Gardete and Antill (2019)	Our Paper (2020)
Search Category	Hotels (unknown)	Video Game (eBay)	Hotels (Travelocity)	Moisturizer (Cosmetic online store)	Hotels (unknown)	Restaurants (Chinese version of Yelp)	Used cars (Shfit)	Multiple categories (Alibaba)
Size	1961 listings 495 users 15days	270 listings 12059 sessions 2 months	2117 listings ~1M sessions 3 months	3577 users 12 months	29065 users 4 months	5465 users 4 months	24116 users 8 months	7 million users 2 years
PC/Mobile Data	PC	PC	PC	PC	PC	PC	PC	Mobile
Transactions	Yes	Yes	Yes	Yes	Yes	Partial	Partial	Yes
Search path	Yes	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Search refinements	Yes	No	Yes	No	Yes	No	Yes	Yes
Search queries	No	Partial	No	No	No	No	No	Yes
Search engagement (time on searching /clicking)	No	No	No	No	No	Yes	Yes	Yes
Consumer panel	No	No	No	Yes	Yes	Yes	Yes	Yes
Search purpose	Product fit	Price	Product fit	Product fit	Product fit	Product fit	Product fit	Flexible (goal-directed or exploratory search)

Table 2 Summary Statistics of the Treatment Group and the Control Group

Category	Smart Trash Can (Treatment)			Vacuum Robot(Control)		
	Before	After	Difference	Before	After	Difference
	(1)	(2)	(3)	(4)	(5)	(6)
Panel A: Matching Outcomes						
1(#Click>0)	0.0242	0.0392	0.1500***	0.0124	0.0151	0.027***
	(0.1535)	(0.1940)	[35.95]	(0.1106)	(0.1219)	[28.40]
1(#Buy>0)	0.0009	0.0014	0.0005***	0.0004	0.0005	0.0001***
	(0.0302)	(0.0383)	[6.81]	(0.0188)	(0.0221)	[8.21]
Panel B: Search Intensity						
#Listing	47.66	42.25	-5.41***	46.72	46.08	0.64
Viewed | #Click>0	(54.66)	(47.24)	[-6.63]	(63.22)	(57.38)	[-1.39]
#Listing	3.58	3.22	-0.36***	2.74	2.81	0.07
Clicked | #Click>0	(0.86)	(1.02)	[-6.29]	(0.49)	(0.53)	[1.27]
Clicking	122.09	108.94	-13.15***	103.68	101.11	2.57**
Time(s) | #Click>0	(34.82)	(37.95)	[-5.44]	(24.60)	(23.57)	[2.05]
#Observations	376670	336334		3204014	2881557	

Note: Column (1)-(2) and (4)-(5) show means with standard deviations in parentheses. Column (3) and (6) show the p-value of t-test for the differences with t-statistic in the brackets. * indicates $p < 0.10$, ** indicates $p < 0.05$, and *** indicates $p < 0.01$

Table 3: The Effect of Category Refinement on Matching Outcomes

Model	DID		Triple Differences	
Dependent Variable	1(#click>0)	1(#purchase>0)	1(#click>0)	1(#purchase>0)
After × Smart Trash Can	0.0069***	0.00026***		
	(0.0013)	(0.00005)		
After × Trash Can			-0.0015	-0.00003
			(0.0008)	(0.00003)
After × Trash Can × Smart Trash Can			0.0083***	0.00028***
			(0.0009)	(0.00005)
Mean of Dependent Variable	0.0185	0.00072	0.0185	0.00072
Query Fixed Effect	Yes	Yes	Yes	Yes
Week of Month Fixed Effect	Yes	Yes	Yes	Yes
Adjusted R^2	0.0281	0.0011	0.0241	0.0008
Observations	7093315	7093315	14011400	14011400

Note: Table 3 reports the estimates of the difference-in-differences (DID) model and the triple differences model, where the first dependent variable is 1 if a consumer makes at least one click during a search session and 0 otherwise, and the second dependent variable is 1 if a consumer makes at least one purchase during a search session and 0 otherwise. We control for consumers' previous search and purchase behaviors in related categories and characteristics of search listings such as product prices. We include search query fixed effects and week of month fixed effects. Standard errors (in parentheses) are robust to heteroscedasticity and clustered by query. * indicates $p < 0.10$, ** indicates $p < 0.05$, and *** indicates $p < 0.01$

Table 4: The Effect of Category Refinement on Search Intensity

Model	DID			Triple Differences		
Dependent Variable	#Listings Viewed |#click>0	#Listings Clicked |#click>0	Clicking Time |#click>0	#Listings Viewed |#click>0	#Listings Clicked |#click>0	Clicking Time |#click>0
After × Smart Trash Can	-0.0442**	-0.0377***	-0.0661***			
	(0.0170)	(0.0083)	(0.0179)			
After ×Trash Can				-0.0340***	-0.0094	-0.0253
				(0.0099)	(0.0074)	(0.0160)
After × Trash Can × Smart Trash Can				-0.0081	-0.0281**	-0.0405*
				(0.0125)	(0.0094)	(0.0173)
Mean of Dependent Variable	47.66	3.58	122.09	47.66	3.58	122.09
Query Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
Week of Month Fixed Effect	Yes	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.1046	0.0445	0.0338	0.1751	0.07288	0.0328
Observations	105464	105464	105464	159697	159697	159697

Note: Table 4 reports the estimates of the DID model and the triple differences model with another set of dependent variables depicting consumer search intensity. We restrict our sample to consumers who make at least one click during a search session. The first dependent variable is number of listings a consumer views during a search session. The other two are number of listings a consumer clicks and the total time she spends on these clicked listings. We take logs for all these three dependent variables and use the same specifications as we did in Table 3. Standard errors (in parentheses) are robust to heteroscedasticity and clustered by query. * indicates $p < 0.10$, ** indicates $p < 0.05$, and *** indicates $p < 0.01$

Table 5: The Effect of Category Refinement on Consumers with Heterogeneous Shopping Needs

	Matching Outcomes		Search Intensity		
Dependent Variable	1(#click>0)	1(#purchase>0)	#Listings Viewed | #click>0	#Listings Clicked | #click>0	Clicking Time | #click>0
After × Smart Trash Can × General Interest	0.0010*** (0.0003)	0.00035*** (0.00006)	-0.0433* (0.0182)	-0.0371*** (0.0083)	-0.0619*** (0.0179)
After × Smart Trash Can × Specific Needs	0.0316*** (0.0015)	0.00041 (0.00028)	-0.0528 (0.0354)	-0.0437 (0.0321)	-0.1063 (0.0570)
Query Fixed Effect	Yes	Yes	Yes	Yes	Yes
Week of Month Fixed Effect	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.02845	0.0011	0.1046	0.0411	0.0338
Observations	6798575	6798575	105464	105464	105464

Note: Table 5 reports the estimates of the DID model with two sets of dependent variables we previously used in Table 3 and Table 4. We interact *After* × *Smart Trash Can* with two dummies to allow for consumers' heterogeneous shopping needs. The dummy "Specific Needs" is 1 if the search query contains specific brand names or application scenarios and 0 otherwise. The dummy "General Interest" is 1 if the search query only includes category names and 0 otherwise. * indicates $p < 0.10$, ** indicates $p < 0.05$, and *** indicates $p < 0.01$

Table 6: The Long-Run Effect of Category Refinement on Consumer Engagement with the Platform

Dependent variables	#Searches consumers conduct in the following week	#Search queries consumer use in the following week	# Days consumers log into the platform in the following week
After x Small Trash Can × Goal-directed Searchers	0.0324** 0.0124	0.0367** 0.0138	0.0157*** 0.0026
After × Smart Trash Can × Exploratory Searchers	-0.0547* 0.0263	-0.0559* 0.0263	-0.0414*** 0.0064
Query fixed effects	Yes	Yes	Yes
Week of month fixed effects	Yes	Yes	Yes
Mean of Dependent Variable	26.95	25.92	5.079
Adjusted R^2	0.0200	0.0206	0.0078
Observation	58277	58277	585277

Note: Table 6 reports the estimates of the DID model with the level of analysis at the week level. Three dependent variables measure consumer search and engagement activities in the following week after seeing the more precise search results. We take logs for all these three dependent variables. Standard errors (in parentheses) are robust to heteroscedasticity and clustered by query. * indicates p<0.10, **indicates p<0.05, and *** indicates p<0.01

Table 7: The Long-Run Effect of Category Refinement on Unplanned Purchases on the Platform

Dependent variables	Total amount of purchase in the following week within the relevant categories	Total amount of purchase in the following week within the same second-tier categories	Total amount of purchase in the following week within the same first-tier categories
After × Small Trash Can × Goal-directed Searchers	0.0027***	0.0014***	0.0008**
	(0.0006)	(0.0005)	(0.0003)
After × Smart Trash Can × Exploratory Searchers	-0.0219***	-0.0125***	-0.0092***
	(0.0058)	(0.0046)	(0.0028)
Query fixed effects	Yes	Yes	Yes
Week of month fixed effects	Yes	Yes	Yes
Mean of Dependent Variable	110.15	81.57	86.38
Adjusted R^2	0.0222	0.0184	0.0198
Observation	16833	16833	16833

Note: Table 7 reports the estimates of the DID model with the level of analysis at the week level. Three dependent variables measure consumers' unplanned purchases in categories other than the category "Smart Trash Can" in the following week after seeing the more precise search results. We take logs for all these three dependent variables. Standard errors (in parentheses) are robust to heteroscedasticity and clustered by query. * indicates p<0.10, **indicates p<0.05, and *** indicates p<0.01

Figure 1: An Example of Category Hierarchy on E-commerce Platforms

Figure 2(a): Category Selection on Alibaba

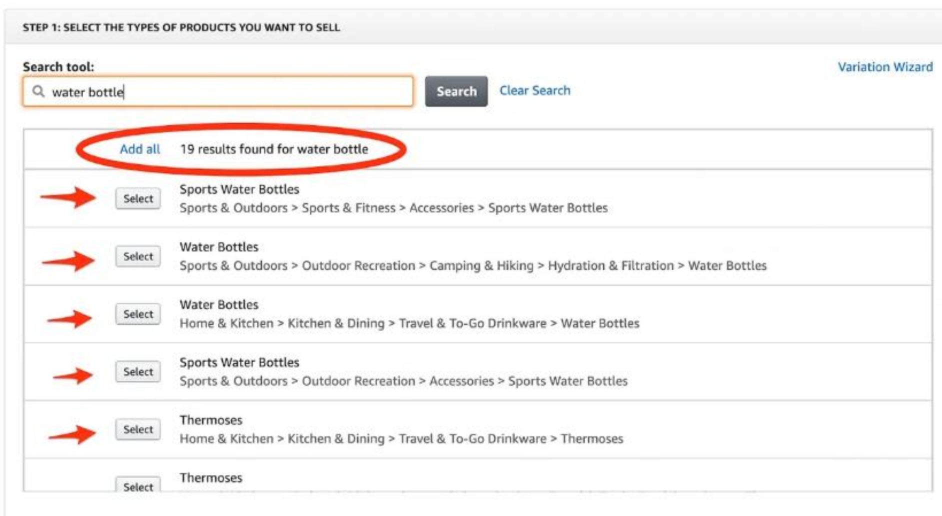

Figure 2(b): Category Selection on Amazon

Note: Figure 2 shows that sellers have to choose a specific category to list their products on e-commerce platforms.

“Smart Trash Can”

The first page of the search result		
Search Query	Smart Trash Can	
	Before the category refinement	After the category refinement
Item 1	General Trash Can	Smart Trash Can
Item 2	General Trash Can	Smart Trash Can
Item 3	General Trash Can	Smart Trash Can
Item 4	General Trash Can	General Trash Can
Item 5	General Trash Can	General Trash Can
Item 6	Smart Trash Can	Smart Trash Can

Figure 3: An Illustration of What Consumers See on Their Mobile Phones before and after the Category Refinement

Note: In Figure 3, we can imagine each column of the right table to be what consumers actually see in their mobile phones before and after the category refinement respectively.

Search Click Purchase

Figure 4: A Typical Online Shopping Journey on Alibaba

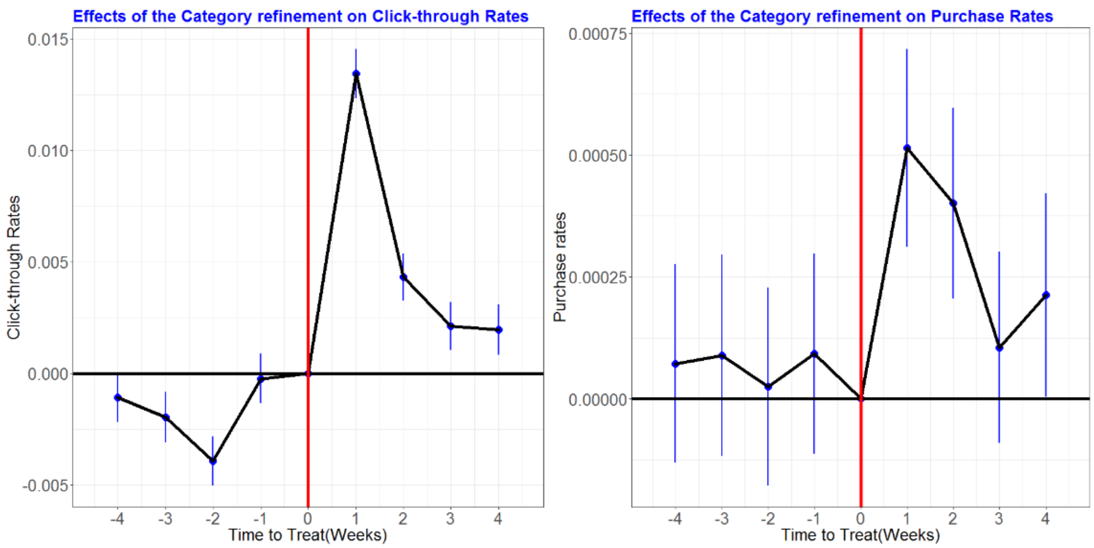

Figure 5 An Event Study of the Effect of Category Refinement on Matching Outcomes

Note: Figure 5 shows DID coefficients and 95 percent confidence intervals from estimation of equation (2) on indicators for category refinement. Standard errors clustered by search query. Includes search query fixed effects and week of month fixed effects. Treatment group defined as search sessions related to the category "Smart Trash Can" and control group defined as search sessions related to the category "Robot Vacuum." Red vertical line represents time of treatment.

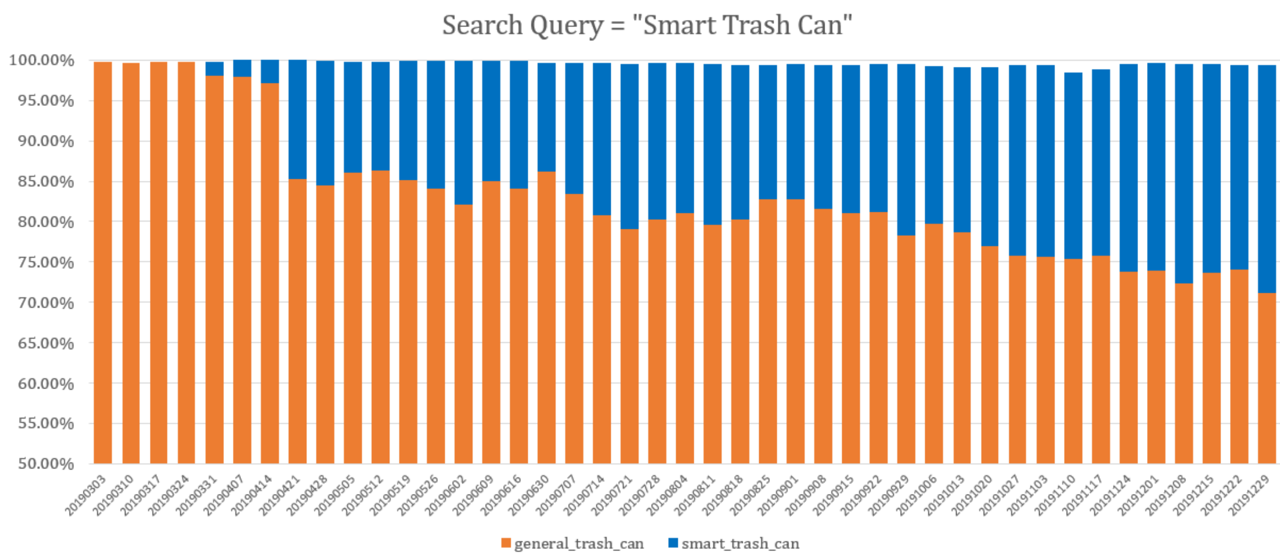

Figure 6: An Example of Targeted Search Traffic Allocation

Note: Figure 6 depicts the distribution ratios of weekly search traffic between two categories for the search query "Smart Trash Can". The Sample period is from 03/13/2019 to 12/29/2019. The category refinement took effect at 04/18/2019. The orange bar represents the ratio of search traffic allocated to the products listed under the category "General Trash Can" as a whole. The blue bar represents the ratio of search traffic allocated to the products belong to the category "Smart Trash Can" as a whole.

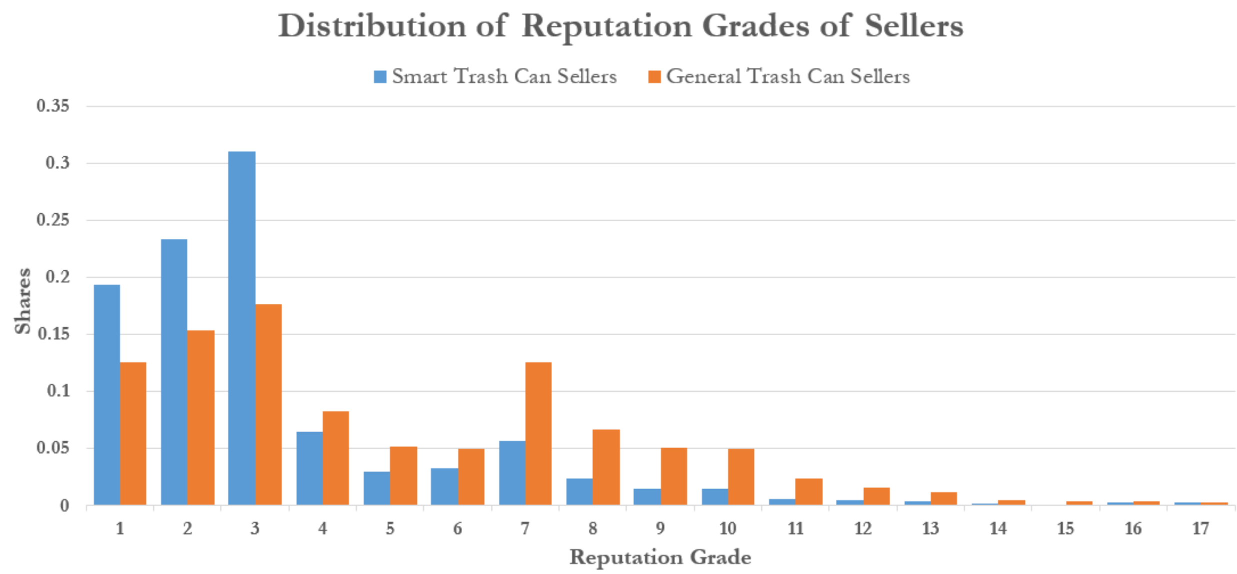

Figure 7: Compare the Distribution of Reputation Grades by Groups

Note: In Figure 7, We pull out all sellers who show up in the search results related to "Smart Trash Can" and divide them into two groups: sellers who enter the new category "Smart Trash Can" and sellers who list their products under the category "General Trash Can". We then compare the reputation grades of these two groups of sellers.

Appendix: Additional Tables and Figures

Table A1: A Partial List of Category Refinement

Effective Date	Original Category	New Category
20190221	E-cigarette/E-liquid	E-cigarette accessories E-liquid E-cigarette
	Candy	Wedding Candy Other Candies
20190318	Other Vegetables	Red pepper Broccoli Other fruits
20190417	Tissue	Paper towel Napkin Tissue
	Trash Can	**Smart Trash Can** **General Trash Cans**
20190517	Smartglasses /VR Device	MR Device AR Device Smartglasses/VR Device

Table A2 Characteristics of Consumers Who Search for Smart Trash Cans before and after the Category Refinement

Characteristics	Before the Refinement (3/18-4/17)	After the Refinement (4/18-5/18)	Difference (p-value)
1(Is a High End User?)	0.1889	0.1936	0.0047
	(0.3915)	(0.3951)	0.65
1(Is a New User?)	0.0051	0.0055	0.0004
	(0.071)	(0.074)	0.09
# Days the consumer logged into	5.85	5.86	0.01
the platform in the last week	(1.60)	(1.58)	0.15
# Days the consumer logged into	23.80	23.72	0.08
the platform in the last month	(6.86)	(6.85)	0.7633
Observations	380163	412829	

Notes: For the first two columns, standard errors are in parentheses. For the third column, p-values are in parentheses. 1(Is a High End User?) is a dummy variable that indicates whether a consumer is labeled as a high end user given her purchase records in the last year. High end users have large purchase power and are price insensitive. 1(Is a New User?) is a dummy variable that indicates whether a consumer just registered her account in the last month.

Table A3: Examples of Two Groups of Search Queries

Query Group	Search Query	Weekly Number of Consumers	Weekly Number of Products
General Interests	Smart Trash Can	2136	12735
	Sensor Motion Trash Can	1476	10069
	Automatic Trash Can	817	8051
	…		
Specific Needs	Trash Can + EKO	151	609
	Smart Trash Can+Automatic Pakaging+Perfect for Home	139	2771
	Trash Can+ Automatic Packaging	95	2797
	…		

Note: Column 3 in Table A4 indicates weekly number of consumers who search for the query. Accordingly, column 4 is the weekly number of products that have ever shown up in the search results for the query.

Table A4 Robustness Checks I: Different control groups

Dependent Variable	Matching Outcomes		Search Intensity		
	1(#click>0)	1(#purchase>0)	#Listings Viewed | #click>0	#Listings Clicked | #click>0	Clicking Time | #click>0
Panel A: Smart Trash Can vs Vacuum					
Smart Trash Can x After	0.0069***	0.00021***	-0.0462***	-0.0245**	-0.0287
	0.0003	0.00006	0.0119	0.0090	0.0190
Query Fixed Effect	Yes	Yes	Yes	Yes	Yes
Week of Month Fixed Effect	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.0373	0.0022	0.0765	0.0284	0.0220
Observation	5412094	5412094	82959	82959	82959
Panel B: Smart Trash Can vs Mop					
Smart Trash Can x After	0.0086***	0.00028***	-0.0453***	-0.0207*	-0.0516*
	0.0002	0.00005	0.0129	0.0099	0.0204
Query Fixed Effect	Yes	Yes	Yes	Yes	Yes
Week of Month Fixed Effect	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.0326	0.0012	0.0724	0.0281	0.0212
Observation	10952676	10952676	64156	64156	64156
Panel C: Smart Trash Can vs Air Purifier					
Smart Trash Can x After	0.0088***	0.00037***	-0.0250*	-0.0132**	-0.0142
	0.0003	0.00006	0.0110	0.0044	0.0204
Query Fixed Effect	Yes	Yes	Yes	Yes	Yes
Week of Month Fixed Effect	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.0330	0.0016	0.1305	0.0474	0.0457
Observation	3444954	3444954	61656	61656	61656

Table A5 Robustness Checks II: Different Samples

Dependent Variable	Matching Outcomes		Search Intensity		
	1(#click>0)	1(#purchase>0)	#Listings Viewed | #click>0	#Listings Clicked | #click>0	Clicking Time | #click>0
Panel A: Long Time Window					
Smart Trash Can x After	0.0125***	0.00038***	-0.0370***	-0.0308***	-0.0486***
	0.0019	0.00004	0.0084	0.0062	0.0141
Query Fixed Effect	Yes	Yes	Yes	Yes	Yes
Week of Month Fixed Effect	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.0274	0.0013	0.1067	0.04078	0.0342
Observation	14472733	14472733	214866	214866	214866
Panel B: Last Year Sample					
Smart Trash Can x After	-0.0011	-0.00007	0.0655**	0.0381*	0.0028
	0.0012	0.00008	0.0214	0.0152	0.0355
Query Fixed Effect	Yes	Yes	Yes	Yes	Yes
Week of Month Fixed Effect	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.0254	0.0011	0.1239	0.0453	0.0323
Observation	5170520	5170520	68494	68494	68494
Panel C: Only Common Users					
Smart Trash Can x After	0.0044***	0.00026***	-0.0329***	-0.0318***	-0.0480***
	0.0015	0.00005	0.0070	0.0072	0.0054
Query Fixed Effect	Yes	Yes	Yes	Yes	Yes
Week of Month Fixed Effect	Yes	Yes	Yes	Yes	Yes
Adjusted R^2	0.0297	0.0042	0.0897	0.0454	0.0368
Observation	186034	186034	4633	4633	4633

Table A6: Summary Statistics on Consumer Search Activities

	Mean	Std.Dev	Median	1st Quantile	3rd quantile	Min	Max
Average Time on Clicks	96.526	110.941	75.046	48.862	114.601	0.046	7274.418
Average Searched Pages	2.967	2.144	2.550	1.788	3.556	1	92
Average Clicks	3.094	1.735	2.730	2	3.667	1	40

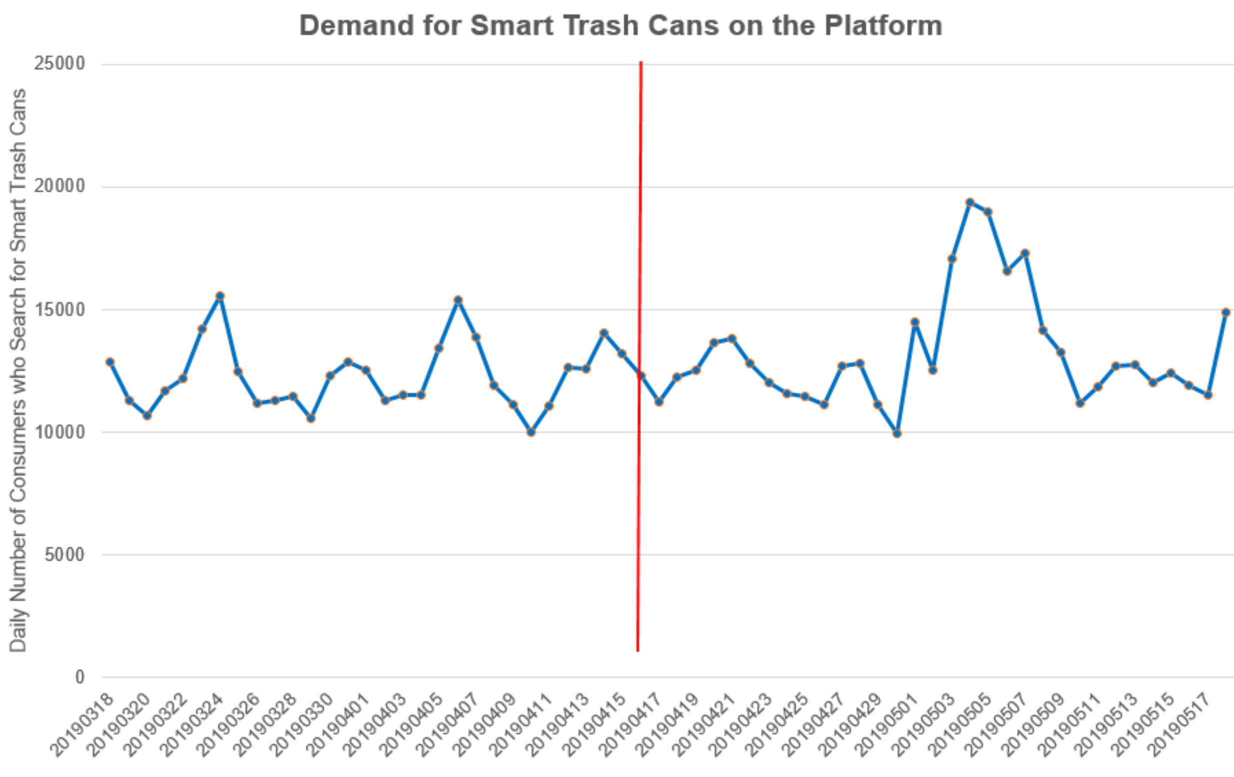

Figure A1: Demand for Smart Trash Cans on the Platform throughout the Sample Period

Note: We calculate the daily number of consumers who search for smart trash cans on the platform. The red line indicates when the category refinement takes effect.