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

Lu Fang
Alibaba Group
fl_fanglu@163.com

(This draft: January 2021)

[1] We thank Hidehiko Ichimura and Ashley Langer for their thoughtful comments. We are grateful to participants at the Symposium on Statistical Challenge in Electronic Commerce Research 2020, the ISMS Marketing Science Conference 2020, the OCIS Doctoral Consortium at the Academy of Management 2020 Annual Meeting, and the 2020 INFORMS Conference on Information Systems and Technology for their comments and suggestions. We also thank Chunxiao Li, Zhe Yuan, Kaifu Zhang, and the economist team at the Alibaba Group for providing instrumental 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 immediate improvement in matching outcomes, however, comes at the cost of a substantial decrease in consumer engagement and unplanned purchases in a longer time horizon for consumers prone to spending more 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 week after the search precision increases. 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

Search engines play a critical role in e-commerce platforms. Consumers arrive at e-commerce platforms with different needs, wants, and demands. Sometimes they seek a highly specific item, such as "Fitbit Versa 2 special edition;" at other times, they come with a vague idea, such as "smart watch." Once consumers enter these ideas as keywords for the platform's search engine, it is the search engine that guides them to various product webpages with the hope of satisfying the consumers' needs and generating revenues for the platform.

There are important tradeoffs that search engine designers must consider when it comes to search traffic. Should the platform just show Fitbit Versa 2 with different colors and add-ons? Or should Fitbit Charge 4 be presented? What about a brand-name competitor, Garmin Forerunner 235, which comes with a higher price tag, or Xiaomi Mi Band, a relatively obscure but up and coming alternative? In this example, the degree of search precision is an important consideration. Higher precision may deliver immediate efficacy, but at the same time, it discourages consumers from exploring alternative options, new products, or relevant categories. In other words, search engines that produce higher precision may also risk losing valuable opportunities for cross-selling. Finding the sweet spot of balancing immediate gratification with long-term consumer engagement has spurred research interest in information systems, marketing, economics, and computer science (Ghose, Ipeirotis, and Li, 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.

More specifically, 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 search redesign changes more than just the consumers' consideration set of products. It may also change the relative salience of different features of the search results or elicit sellers' strategic responses in product presentation, pricing, or other dimensions.[1] Such information is

---

[1] 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.

often unavailable in the extant research. Last but not least, researchers rarely have access to individual-level search panel data that track each consumer across multiple search sessions over time or across multiple categories of products. Because of such restrictions, prior studies often had to assume that each search session represents a different consumer (Ghose, Ipeirotis, and Li, 2019, p.1382; Dinerstein et al. 2018), and each consumer can make, at most, one purchase (Ghose, Ipeirotis, and Li, 2014, p.1653).

To address these challenges, we take advantage of a detailed click-stream dataset with quasi-experimental variation in the precision of search results. This dataset 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 and part of Alibaba group.[2] 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 see 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 was 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 years ending in December 2019. Two advantages of our dataset help overcome the limitations of 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 (Blake, Nosko, and Tadelis, 2016). We overcome this limitation

[2] 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 the heart of its core business, which is to facilitate transactions between sellers and buyers. This also provides us an ideal environment to observe how buyers and sellers adapt to the advance in search technology.

by using consumer search data from mobile transactions, which account for over 95% of transactions made on Taobao during our sample period.[3] Consumers are automatically logged in when starting a search session on their mobile phones. Thus, we can track multiple sessions across time conducted by the same consumer.[4] We are able to link the search sessions of each consumer from the first query all the way through to either a purchase or abandonment of the search. We know not only the number of 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 or books (Ghose, Ipeirotis, and Li, 2014; 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.

For the purpose of our analysis, we choose to focus on the "Trash Can" product category and its related categories of home and personal cleaning supplies, such as mops, trash bags, vacuum cleaners, air purifiers, and storage racks.[5] We pick trash cans because they are well-defined, typical household items. A trash can is a necessity in a household, has no easy substitute, has steady demand throughout the year, and does not evoke too much emotional response. It is a product that 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—the generic "Trash Can" and "Smart Trash Can"[6]—and this category refinement defines our

[3] Taobao's search team confirms that the majority of consumers conducted searches and made transactions on the Taobao App during our sample period (see §4.2 details). Therefore, the cross-channel substitution is less of a concern in our setting (Brynjolfsson, Hu, and Rahman, 2009; Xu et al, 2017).

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

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

[6] Smart trash cans use advanced infrared sensor technology to open the lid automatically when consumers approach the bin and to close it when consumers walk away.

treatment group as consumers whose search queries were related to "Smart Trash Can" or "Trash Can."[7] We identify the control group by looking into the search records of consumers in the treatment group. "Robot Vacuum," "Air Purifier," "Vacuum," and "Mop" were the top alternative keywords that these consumers also searched for. We use consumers who searched "Robot Vacuum" as the control group in our main analysis and consumers who searched "Air Purifier," "Vacuum," and "Mop" as alternative control groups for robustness. We first document the search precision improvement after the category refinement in the treatment group. We then employ a flexible difference-in-differences, and sometimes a triple differences, regression design[8] 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 consumers' activities on the platform in the week following their initial search to measure the category refinement effect on consumer engagement and cross-selling in related categories.

We report four sets of main results addressing (1) matching outcomes, (2) search intensity within the same search session, (3) consumer heterogeneity, and (4) consumer engagement one week after the initial search. First, we find that the matching between consumers and products significantly increased with search precision improvement. Specifically, the average click-through rate of consumers who searched for smart trash cans increased by 37.3% after the category refinement, relative to consumers who searched for unaffected categories. The purchase rate of smart trash cans also increased by roughly 36.1%, resulting in a 64.4% increase in 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 searched for generic trash cans. We discover that this differential effect is driven by both Taobao's search traffic allocation and a surge of entry of smart trash can listings right after the category refinement. While the search refinement implementation was instantaneous, sellers gradually reacted by changing the keywords associated with their products. In the half-year period after the category refinement, the

[7] Taobao started to gradually roll out 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] Consumers who search for "smart trash can" and "trash can" experienced different treatment intensity, as we will explain in Section 5.2.

proportion of search traffic going to the "Smart Trash Can" category increased from 15% to 25%, boosting the weekly number of smart trash can listings on the platform by 150%. As smart trash cans are niche products that cater to a small proportion of consumers, those who searched for smart trash cans would see only a few listings of smart trash cans among many more generic trash cans before the category refinement. The results page for consumers who searched for generic trash cans did not change much with the refinement. Therefore, the category refinement effect on matching outcomes was demonstrated only for consumers searching for smart trash cans.

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

Third, we show that the gain for consumers who searched for smart trash cans also depends on consumer heterogeneity, as indicated by their search queries.[9] Specifically, we categorize two sub-groups of consumers: one group has specific shopping needs and knows exactly which products to buy, such as "Smart Trash Can Xiaomi 7L White," while the other group has a sub-category of products in mind, such as "Motion Sensor Trash Can," but is open to the array of options in this sub-category. We find that, although both groups of consumers seemed to benefit from higher search precision, they gained in different ways. For consumers with specific shopping needs, the gain was just an increase in matching outcomes without a reduction in search intensity. We suspect that this group of consumers tended not to search much before the category refinement. The real efficiency gain was to consumers who had only a sub-category of products in mind. On average, these

[9] Web queries can be grouped into three categories based on users' intent: transactional, informational, and navigational (Broder, 2002; Rose and Levinson, 2004; Jansen et al., 2008). Most search queries in e-commerce platforms belong to the first two categories. Specifically, online shoppers use transactional queries to express specific shopping needs, while informational queries are likely to be imprecise and general.

consumers' click-through rate increased by 20.5%; their purchase rate increased by 31.9%; and they viewed 4.3% fewer listings, clicked into 3.7% fewer listings, and spent 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, as well as 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 catering precisely to her intention; then, the consumer purchases quickly, without much fuss or hassle. Does this gain in efficiency come at some cost? Is this the shopping environment that e-commerce platforms want to create?

To answer these questions, we must consider the fact that an e-commerce platform is a multi-product and multi-category (virtual) shopping mall. Demand spillover, cross-selling, or achieving economies of scope is essential for its business model (Basker, Klimek, and Hoang Van, 2012; Hwang and Park, 2016; Li and Agarwal, 2017; 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 the category refinement.[10] The "goal-directed" consumers care about the efficiency of gathering information in the search process. The positive experiences of quickly getting what they want make this type of consumer more likely to return to the platform for shopping. However, 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 do not know which particular product has the features they need. After browsing some listings and trying different search queries, they may realize that, instead of a rice cooker, what they actually need is a steamer.

Our fourth result shows that for "exploratory" consumers, the immediate improvement in matching outcomes was accompanied by a substantial decrease in consumer engagement and purchases in related categories in a longer time horizon. Too-precise search results may take away the fun of exploration,

[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).

disincentivizing them from revisiting the platform to look through relevant products. For this type of consumer, we find a significant decrease in engagement with the platform in the week following the increase in search precision. The average number of days that they visited the platform during a week decreased by 4.1%; the number of searches in a week decreased by 5.5%; and the number of search queries they used decreased by 5.6%. Consequently, these consumers decreased their spending in the relevant product categories by about 2.2% in the week following the improvement in search precision. As a comparison, although the "goal-directed" searchers seemed to engage more with the platform and make more purchases in relevant categories, the magnitude of this gain was multifold smaller than the effects on the "exploratory" type. The loss in consumer engagement and cross-selling dominated the gain.

Taken together, our four sets of findings illustrate the tradeoff between exploitation and exploration in platform search design. A search engine, which directs search traffic, helps consumers to exploit and explore the search results. More-precise search results can generate immediate efficacy in satisfying consumers' shopping needs and boost the platform's transaction volume, but too much precision can suppress 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 cutoff, as the right amount of search precision depends on product categories, the distribution of consumer preferences and their search habits, and the potential for scope economies.

Our paper makes several contributions. From an empirical point of view, we enrich the empirical literature by bringing in one of the most granular consumer search data over time and across multiple product categories. We use these data to demonstrate the multi-faceted effects of search engine precision improvement. Researchers in computer science and design science have long shown 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). Therefore, platforms, including Taobao, have devoted significant efforts to improving search engine precision through category refinements. However, theorists in economics and marketing (Yang, 2013; Zhong, 2019) suggest that more precise search engines may decrease consumer search intensity. In turn, reduced search intensity could discourage consumers from exploring new products, resulting in missed cross-selling opportunities. For example, Fong (2017) showed, through a field experiment, that targeted email offers reduced consumer search activities online. In the theoretical models by Hagiu and Wright (2020) as well as Rhodes, Watanabe, and Zhou (2020), decreased consumer search activities lead to missed opportunities for cross-selling. These studies have developed in their own domains but have not been integrated thus far, and they all point to an important tradeoff of search engine design that has yet to be empirically documented.

Synthesizing these streams of studies, we leverage our massive, granular consumer search and purchase data to highlight these tradeoffs in search engine design. Although there is a natural urge for search engines to continuously improve their precision, we show the unintended, negative consequences of search precision improvement. We suggest that e-commerce platforms should consider the search- and sale-spillover across multiple categories for consumers who might return for business over the foreseeable future. The platform should balance the short-term gains from increasing search precision with the long-term benefits from encouraging consumer exploration. The exploitation-exploration dilemma is not unique to search design—many research fields, ranging from organizational innovation and depression treatment to reinforcement learning, demonstrate similar conflicts (March, 1991; Currie and MacLeod, 2020; Sutton and Barto, 1998). For that reason, our study has important implications for practice as well, especially in the current age of machine learning and predictive analytics.

We discuss 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 the results in Sections 6 to 8. Section 9 discusses interpretations, implications, and caveats of our results as concluding remarks.

2. Literature Review

Our study contributes to three strands of literature: consumer search, platform search design, and the exploitation-exploration model. In this section, we briefly discuss 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 indicate that substantial search costs can discourage consumers from considering all available products on 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 when consumers search for a good, price is the characteristic of which they are most uncertain (Sorensen, 2000; Hortacsu and Syverson, 2004; Hong and Shum, 2006). The rise of e-commerce has drawn research attention to the more complex process of a multi-attribute search as the detailed web browsing data have become available (Bronnenberg, Kim and Mela, 2016). Recent empirical studies examine consumer search behaviors within a platform when the purpose of a search is to find a good fit (Kim, Albuquerque, and Bronnenberg, 2010; Koulayev, 2014). These papers often consider a single-category setting in which consumers have an exact product in mind, and their purpose in searching is to acquire information and resolve uncertainty about the product. A critical identification assumption is that consumers have rational expectations about the distribution of product attributes before searching (Honka, Hortaçsu, and Wildenbeest, 2019).[12] Thus, a consumer stops searching either because of a high valuation for the product already found, which results in a successful search, or because of a high search cost that discourages consumers from continuing to search.

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] Improving search quality, at

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

[13] A further distinction of search models in the literature is the search method that consumers use when searching. Theoretical papers imply that consumers adopt either a sequential or a simultaneous search model (McCall, 1970; Weitzman, 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, Hortaçsu, and Wildenbeest, 2012; Honka and Chintagunta, 2017).

[14] Yang (2013) incorporates the quality of search into the traditional search model to explain how the widespread usage of the Internet leads to long-tail effects. He finds that a decrease in search costs and an increase in search quality have different

first glance, should improve the performance of search algorithms across the board.[14] Our overall results confirm this big-picture insight, but our long-run results paint a more nuanced picture of tradeoffs. In particular, we find that an increase in search quality of one category can discourage consumers from exploring other related product categories and decrease their engagement with the platform in the long run.

By considering the role of search technologies in shaping search results, our paper is also related to recent studies on the effects of ranking algorithms on consumer choices (Ghose, Ipeirotis, and Li, 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, in the sense that consumers may not know exactly what to buy at the beginning of a search process. Second, we track consumer search activities across multiple categories, which allows 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.

[Insert Table 1 about here.]

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, Hu, and Simester, 2011).[15] There has been a productive effort to optimize search design to increase consumer surplus and boost search engine revenues (Chen and Yao, 2017; Ghose, Ipeirotis, and Li, 2019; Gu and Wang, 2019; Gardete and Hunter, 2019; Zhang et al., 2019; Yoganarasimhan, 2020). Chen and Yao (2017) highlight the value

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.

[14] For example, a recent study by Sun et al. (2020) conducted a large-scale field experiment to investigate the use of personal data, including demographics, and past clicking and purchase behaviors of Taobao's recommendations, a starting point of search for many consumers. They show that banning the use of personal data leads to a substantial decrease in consumer engagement as measured by product views, click-through rate and purchase rate.

[15] By alerting researchers 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 (Einav et al., 2016; Chen and Shelton, 2016; Fradkin, 2017; Horton, 2018; Basu, Bhaskaran, and, Mukherjee, 2019).

of refinement tools in consumers' online search and predict that consumers will search less and have lower utilities when sorting or filter options are not available. Gu and Wang (2019) 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, Ipeirotis, and Li (2019) show that platforms can improve consumer online search experience by incorporating social content on the summary page of search results. Most of these papers focus on the short-run effects of platform design on consumer search and purchase behaviors.

Our unique data allow us to contribute to this literature by analyzing the effects of platform design on consumer behaviors both in the short run and in the long run. Our study highlights the possibility that the long-term positive impact is not guaranteed even when there is a short-term benefit. In addition, we also examine how sellers respond to the changes in platform design in the long run, especially in terms of their entry decisions. Our results indicate that niche products gain more market shares, and the distribution of sales becomes less concentrated as search precision improves. More broadly, our paper is also related to the literature on product design and the long-tail effect in online markets (Kuksov, 2004; Brynjolfsson, Hu, and Simester, 2011; Bar-Isaac, Caruana, and Gunat, 2012; Yang, 2013; Larson, 2013).

2.3 Exploitation and Exploration

Our study also builds on the broad exploitation and exploration literature in computer science, statistics, organization science, and economics (Schumpeter, 1934; Holland, 1975; Kuran, 1988; March, 1991; Sutton and Barto, 1998; Benner and Tushman, 2003; Dewan and Ren, 2007). On the one hand, 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 with 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 doctors' 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; Wang, Li, and Luo, 2019).[16] Although diverse in research topics, these studies can be unified in the multi-armed bandit framework in which 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 (Johar, Mookerjee, and Sarkar, 2014; Dzyabura and Hauser 2019).

Our paper extends the literature on exploitation and exploration to the study of platform search design because our dataset allows us to examine how consumers search not only within a particular category but also across categories. The tradeoff between exploitation and exploration arises because of an organization's limited resources. Similarly, consumers have limited attention during their search process, and, accordingly, any search engine design should accommodate the fact that e-commerce platforms must prioritize what information to show and the sequence of the search results displayed. If search engines prioritize or even focus solely on results in a best-guess category, they are likely to yield desirable results for the consumers within that category. At the same time, the search engine will have to sacrifice the opportunity for cross-selling or bringing consumers' attention to products in a different category. To the best of our knowledge, our paper is among the first in the literature to empirically quantify the tradeoff between exploitation and exploration for search engines.

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 (Arazy and Woo, 2007; Shen, Ruvini, and Sarwar, 2012; Roshdi and Roohparvar, 2015).[17] Figure 1 provides an example of the category hierarchy on e-commerce platforms. "Sports" is in the first-tier category, under which there are four second-tier categories:

[16] Studies on recommendation systems have argued that predictive accuracy should not be the only focus of recommendation algorithms (McNee, Riedl, 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; Zhang et al, 2020).

[17] For example, one of the earliest success stories of the Internet is Yahoo, which essentially tried to create a catalog of the Internet. http://misc.library.ucsb.edu/untangle/callery.html

"Sports shoes," "Sports clothes," "Equipment," and "Accessories." The category "Sports Shoes" is further refined into three third-tier subcategories: "Tennis Shoes," "Soccer Shoes," and "Basketball Shoes."

[Insert Figure 1 about here.]

Categorization can not only navigate consumers to the products they want but also organize the relevant products for the search engine to retrieve. Consumers can go through the category menu and search for one of the product categories. Alternatively, they can start a search session by typing a search query into the search box. Once the search query is understood, the search engine then retrieves several product categories that are most likely to match the search query and further restricts its attention to the products that belong to these categories.[18] Finally, the search engine ranks all of these products in a specific order and delivers the results to consumers. In this matching and ranking process, product categorization allows the search engine to better "guess" consumers' shopping needs behind their search queries (Xing et al., 2008; Bilal, 2012).

Without opening physical stores, a critical way for sellers to "position" themselves in this vast two-sided online market is to list their products in a particular category and let the search engine guide consumers toward them. Online search is the communication process between consumers and search engines, whereby 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 consumers' questions 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 that search term in the search box. 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 under which 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. Sellers have to choose one specific category node from the category menu when listing a product for sale. Figure 2b shows a similar process on Amazon.

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

[Insert 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[19] and capturing a 55.9% share of all Chinese e-commerce retail sales.[20] Taobao enables small businesses and individuals to reach consumers and is very similar to eBay in the US. Tmall focuses more on helping large companies and multinational brands such as Nike and Apple to sell directly to consumers. In contrast, Alibaba.com is a business-to-business trading platform for manufacturers and suppliers to find vendors and purchase merchandise in bulk.

Search engines routinely refine and improve their catalog of data to increase search precision (Dumais and Chen, 2000; Xing et al., 2008; Bilal, 2012; Farhoodi, Ghidary, and Yari, 2013). For our identification purposes, we leverage such a routine change on Taobao that occurred in 2019. With an increasing number of products and sellers available on the market, the platform has been refining broad categories into narrow subcategories. Appendix Table A1 provides a partial list of the recategorization on Taobao. For example, before the 2019 refinement, e-cigarettes and e-cigarette accessories belonged to the same leaf category. Then, Taobao divided this leaf category into an "E-cigarette" category and an "E-cigarette accessory" category. As another example, the category "Tissue" was refined into three subcategories: "Paper Towel," "Napkin," and "General Tissue." Note that there is no obvious indication to consumers that a search refinement has taken place, as this occurs completely behind the scenes.

4. Data

4.1 Sample Selection

To estimate the effect of category refinements, we focus on the refinement of one particular category. After April 17, 2019, Taobao divided the category "Trash Can" into two subcategories: "Smart Trash Can" and generic "Trash Can." Smart trash cans use advanced infrared sensor technology to open the lid automatically

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

when consumers approach the bin and close it when consumers walk away. After the category refinement, the search engine gave much more search traffic to the products belonging to the subcategories, especially for consumers who searched for niche products such as smart trash cans.[21] Figure 3 illustrates how the refinement of the category "Trash Can" can affect consumers' search results. We can imagine each column of the table on the right side of the figure to be what consumers actually see on their mobile phones. Before the category refinement, consumers who submitted the query "Smart Trash Can" would see mostly generic trash cans in the search results. With the category "Trash Can" refined into the two subcategories, "Smart Trash Can" and generic "Trash Can," consumers would find many more smart trash cans in the search results.

[Insert Figure 3 about here.]

We choose this product category for two reasons. First, the demand pattern for this product category shows little seasonality.[22] Other product categories on the platform exhibit a time trend of demand, complicating our estimation of the effect of platform search design.[23] 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 generic trash cans. Thus, the category refinement should significantly improve the matching outcomes of smart trash cans but should have an insignificant effect on generic trash cans. With one heavily treated group (smart trash can) and one slightly treated group (generic trash can), we can control for potential confounders that may influence the search outcomes.

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 select all consumers who have used these search queries from the universe of consumers on the platform. For each consumer, we keep all search results derived only from the consumer's first search query in the first search session that is related to the product category.

[21] In Section 7.1, we provide empirical support for this argument to show the underlying mechanism of category refinements.
[22] Appendix Figure A1 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 steady throughout our sample period.
[23] For example, seasonal fruits such as watermelon demonstrate stronger demand during the summer.

The idea is to capture the very beginning of a consumer's search process and to examine how the search precision in the initial search session impacts her exploration efforts in subsequent searches.

We use consumers whose search queries were related to the category "Smart Trash Can" as the treatment group. We identify the control group by looking into the search records of consumers in the treatment group. "Robot Vacuum," "Air Purifier," "Vacuum" and "Mop" were the top alternative keywords that these consumers also searched for. Although we have access to all categories of products, we use consumers whose search queries were related to "Robot Vacuum" as the control group in our main analysis. We also use other search queries in our robustness checks.[24] We compare the changes in consumers' search and purchase behaviors in the treatment group before and after the category refinement, relative to the same changes over time among consumers in the control group.[25]

4.2 Summary Statistics

We use administrative mobile search data from Alibaba. Our dataset includes around 7 million consumers over two years ending in December 2019. Mobile transactions account for over 95% of total transactions on the platform during our sample period. Our search data paint a comprehensive picture of consumers' footprints when they use the shopping App. The shopping experience on e-commerce platforms can be summarized as search, click, and purchase. Figure 4 illustrates these three stages. It starts with consumers describing what they want by typing a query in the search box. Then, the search engine interprets the query and returns the relevant products to them. Next, consumers look through the list of products in the search results and click on some products that match their preferences to get further information. After comparing the pros and cons of every item on which they have clicked, consumers may end up buying the one that they like the most. Our data capture what consumers see and every decision they make throughout this journey.

[Insert Figure 4 about here.]

[24] See Section 6.4 for details. We use the category "Air Purifier," "Vacuum" and "Mop" as the alternative control group.
[25] In Appendix Table A2, 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 category "Smart Trash Can" before and after the category refinement.

Our main analysis uses the search and purchase data four weeks before and after the refinement took place.[26] The unit observation is at the search-session level. Each search session is defined by a unique user ID, a search query, and a specific date.[27] In a search session, we observe how many listings a consumer views 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 the amount of time she spends on all the listings that she clicks on. Based on whether a consumer makes any clicks or any purchases after a search, we further generate two dummies to summarize the outcomes of a search session. At the consumer level, we can observe each consumer's entire search and purchase histories throughout our sample period.[28] We know whether a consumer logs into the platform and has any engagement with it. Thus, we can track consumers' search and purchase activities in subsequent sessions, after the improvement upon the search precision of the initial search session. At the seller level, we observe a detailed transaction history of every item a seller ever makes available for sale. Therefore, we can directly calculate the transaction revenues that a seller earns from each item. We also know each seller's characteristics, such as ratings and the 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 consumers' search outcomes and intensity 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 of the click-through rate and purchase rate in the category "Smart Trash Can", could be a consequence of the platform category refinement. In addition, consumers who search for smart trash cans click, on average, fewer times before making a purchase after the category refinement, while there is no significant change in search intensity for consumers who search for robot vacuums.

[Insert Table 2 about here.]

[26]We lengthen the time window to eight weeks before and after the category refinement in our robustness checks.
[27] 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.
[28] All consumers are anonymous in our data. We do not use any personal information of consumers.

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 or purchase decisions for consumer i, who searches using 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 March 18, 2019 to May 18, 2019, and *After* equals one for April 18, 2019 through May 18, 2019. X_{iqt} is a vector of covariates that capture consumer i's previous search or purchase behaviors. 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 day t. We also control for the average price of the search listings that consumer i sees when she enters query q on day t. In addition, 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, Duflo, and Mullainathan (2004), we cluster the standard error at the query level to allow for the 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" with 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 should show similar trends to the treated group (Card and Krueger, 1994; Angrist and Pischke, 2008). To explore the validity of the design, we conduct an "event time" analysis. This allows an examination of the pre-trends (Autor, 2003; Chetty et al., 2014). We replace *After* $\times$ *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 the generic "Trash Can" as another treatment group. We then use a triple differences analysis (Goldfarb and Tucker, 2011; Rishika et al., 2013) as a placebo test. This new treatment group refers to consumers who type in the keyword of "Trash Can" rather than "Smart Trash Can." The first part of the triple differences analysis is the difference in search outcomes and intensity between the two treatment groups (*Smart Trash Can* and the generic *Trash Can*) and the control group (*Robot Vacuum*) after the category refinement, minus the difference in search outcomes and 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 focuses on the coefficient for the triple interaction term $After_t \times TrashCan_q \times SmartTrashCan_q$. This captures the difference in search outcomes and intensity between the two treatment groups (consumers searching for the niche category "Smart Trash Can" vs. those searching for the generic "Trash Can") after the category refinement, minus the difference in search outcomes and intensity between 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 controlling for those factors, it will remove potential confounders that might lead the category "Trash Can" to be chosen for the treatment (category refinement). For example, suppose Taobao pays special attention to the target category and allocates more marketing expenses to that category. In that case, we should see significant changes in both the generic subcategory "Trash Can" and the subcategory "Smart Trash Can" after the refinement, because they were in the same category prior to the refinement. However, if becoming the target category per se does not affect consumer behavior, the category refinement should significantly improve the matching outcomes of consumers who search for smart trash cans, but should have little effect on consumers who search for generic 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: the generic "Trash Can" and "Smart Trash Can." Our treatment group for the main analysis is consumers who searched for smart trash cans during our sample period, while the control group is those who searched 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 made at least one click after being presented with the search listings. The other one indicates whether a consumer made s at least one purchase during the first search session. These two variables generally summarize whether a consumer was interested in or satisfied with the search results that the search engine delivered based on her search keyword.[29] We expect that consumers were more likely to click and end up with a purchase when the search results matched 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 searched for smart trash cans, relative to those who searched for robot vacuums, were 0.69 percentage points more likely to click after the category refinement than prior to the refinement, translating into a 37.3% increase in the average click-through rate. The purchase rate in the

[29] 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.

treatment group also increased by roughly 36.1% after the category refinement, resulting in a 64.4% increase in gross merchandise volume in the category "Smart Trash Can."[30]

[Insert 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 searched for the generic 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 searched for generic trash cans, neither the click-through rate nor the purchase rate significantly changed after the category refinement. Thus, the category refinement has negligible effects on these consumers, alleviating the concern about selection bias in the category refinement.

To check the validity of the difference-in-differences design, we use an "event time" analysis. This allows us to examine the pre-trends. In Figure 5, we plot the week by $SmartTrashCan_q$ interactions using an 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 dummies smaller than 16, most treatment coefficients are less than 0.005 points in magnitude and seldom reach statistical significance. After the category refinement, click-through rates and purchase rates increased significantly among consumers who searched for smart trash cans.[31]

[Insert 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

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

[31] In Figure 5, the click-through rate and the purchase rate show a slight 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 on and purchase products from sellers with fewer ratings (Ghose, Ipeirotis, and Li, 2014; Dinerstein et al., 2018).

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 viewed after entering a search query. The key decision a consumer made in this viewing process was whether or not to scroll down. If she was particularly interested in one of the listings in the search results, she needed to obtain additional information by clicking on it. In light of this, the other two measures focus on how many listings a consumer clicked on during the viewing process and the total time she spent on these clicked listings. These two measures represent the search on the extensive margin (similar to exploration) and the search on the intensive margin (similar to exploitation), respectively (Ursu, Wang, and Chintagunta, 2019).

Similar to Table 3, Table 4 presents the results from the difference-in-differences model and the triple differences model. We restrict our sample to consumers who made at least one click during a search session. [32] The first column of Table 4 shows results with the dependent variable being the 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 spent on these clicked listings as the dependent variable, respectively. We take logs for all three dependent variables when running regressions.

Results from Table 4 indicate that consumers reduced search intensity after the category refinement. As we can see from the estimates of the difference-in-differences model, for consumers who clicked on listings, the total number of viewed listings decreased by 4.4% after the category refinement. The category refinement also led to consumers clicking on 3.7% fewer listings and spending 6.6% less time on clicking. Our results from the triple differences model also suggest that the effect of category refinement on search intensity was significant mainly for consumers who searched for smart trash cans.[33] Our findings are consistent with the theoretical prediction in Yang (2013). His model predicts that consumers' overall search might decrease if there

[32] 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% of consumers do not click on an item 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.

[33] Column 4 of Table 4 indicates that consumers who search for the "generic" 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 generic 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 generic 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).

were an increase in search quality. As search precision improves, consumers are more likely to find the right product and purchase more quickly during the search process.

[Insert Table 4 about here.]

6.3 Consumer Heterogeneity

To understand how matching outcomes and search intensity vary across different consumers, we estimate the heterogeneous treatment effects of the category refinement. In Table 3 and Table 4, we group together all consumers who search for a product category and obtain 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 in on 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.[34] Appendix Table A3 provides examples of these two groups of search queries. The "general interest" subgroup includes only search queries with the name of a product category or an imprecise description of desirable features of a product (Broder et al., 2002; Rose and Levinson, 2004; Jansen et al., 2008; Du et al., 2017; Yang et al., 2014). For example, "Smart Trash Can" and "Sensor Motion Trash Can" are the two most searched queries in the "Smart Trash Can" category. Out of all consumers in our sample, 30.49% used these two generic queries when searching for smart trash cans. These consumers have just a category of products in mind and are often at the early stage of shopping. They learn and adjust their preferences as they explore more products in the search results. We identify the "specific need" subgroup as consumers whose search queries 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 they want to buy when they enter this kind of query in the search box. They have

[34] Web queries can be grouped into three categories based on users' intent: transactional, informational, and navigational (Broder, 2002; Rose and Levinson, 2004; Jansen et al., 2008). Users use transactional queries to look for specific products (e.g., "buy Apple watch"); in informational queries, users look for information (e.g., "Apple watch features"); navigational queries help users navigate to a website (e.g., "apple.com"). Most search queries in e-commerce platforms belong to the first two categories. Specifically, online shoppers use transactional queries to express specific shopping needs, while informational queries are likely to be imprecise and general.

gathered enough information about smart trash cans and are ready to make a purchase. After grouping all search queries into the 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 the results of this exercise. As for the matching outcomes, the first two columns indicate that the category refinement significantly increased click-through rates for these two subgroups of consumers. This positive matching effect was especially pronounced for consumers with specific shopping needs. After the category refinement, these consumers were 3.2% more likely to make at least one click after viewing the search results.[35] Interestingly, as suggested by the next three columns, consumers with general shopping interests showed significantly lower search intensity as a result of the category refinement. This subgroup of consumers viewed 4.3% fewer listings, made 3.7% fewer clicks, and spent 6.2% less time clicking. However, for consumers with specific shopping needs, we do not find significant changes in their search intensity after the category refinement.

Overall, our results empirically illustrate a trade-off between exploitation and exploration in e-commerce search engine design. As search precision 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 fewer products before landing a deal. The positive matching effect is more pronounced for consumers with specific shopping intents who have already made up their minds, while the negative search deterrence effect is disproportionately high on consumers with general shopping interests who have yet to decide whether to purchase or what to purchase. The latter could represent lost sales opportunities.

[Insert Table 5 about here.]

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

6.4 Robustness Checks

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

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 groups, respectively, while maintaining the category "Smart Trash Can" as the treatment group. We choose these unaffected product categories as candidates for our control groups for the following reasons. First, the characteristics of consumers searching for these product categories are close to those of consumers in our treatment group. We examine the search records of consumers who searched for smart trash cans, and find that the top alternative keywords that these consumers also searched for are "Robot Vacuum," "Air Purifier," "Vacuum," and "Mop." Second, search and purchase outcomes in these product categories have trends similar to those of the treated group during our pre-treatment period, thus meeting the identifying consumption of the difference-in-differences model. We obtain similar estimates of the effects of the category refinement on matching outcomes and search intensity, as we see in Tables 3 and 4.

We further re-estimate our model using alternative samples to address other empirical concerns. First, our findings may be merely a result of seasonal promotions if other marketing efforts are accompanied with the category refinement during our sample period—i.e., from March 18, 2019 to May 18, 2019 . Thus, we estimate the model using the sample from the previous year—i.e., from March 18, 2018 to May 18, 2018. We find that there was no significant change in either click-through rates or purchase rates for consumers who searched for smart trash cans. Hence, our findings are driven mainly by the changes in the search algorithms. In addition, one may worry that consumers who searched for smart trash cans are not comparable to those who searched for the products in the control groups. To address this, we restrict our attention to the consumers who searched for both "Smart Trash Can" and "Robot Vacuum." This generates a "common-user" subsample of around 0.18 million consumers, which allows us to control for unobserved consumer characteristics. Estimating our models on this subsample yields highly consistent findings. Lastly, we estimate a sample of a longer time window of four months, and the results remain consistent. We use the search and purchase data eight weeks before and after the category refinement took place—i.e., from February 18, 2019 to June 18, 2019.

7. The Long-Run Effects of Improving Search Precision

In the main analysis, we documented that increasing search precision helps consumers find matched products more quickly. An unintended but inevitable consequence is that it reduces the consumers' exposure to other products on the platform. For example, consider the consumers' search process in physical stores. When they look for a particular product, they will inevitably be exposed to other products, which, in turn, might generate impromptu purchases (Silley et al., 2010; Hui et al., 2013). For online platforms such as Alibaba, better search engines lead consumers to make a purchase right away, but that also means that they will spend less time on the site and will be less likely to make unplanned purchases. Therefore, a follow-up question is how such a decrease in consumers' search intensity affects their purchase behaviors online. Answers to this question might not be exactly the same as in the offline context because there is a fixed cost for offline shopping (the effort and time of traveling to a location), which may not be true for the online context. [36] Therefore, it is ex-ante unclear whether a better-designed search engine would lead to lower long-term consumer engagement.

7.1 Consumer Engagement

This section examines the impacts of the category refinement on consumer engagement with the platform over a longer period of time. 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 changes after seeing the more-precise search results. We use the number of search sessions and search queries to describe consumer search activities in the week following the initial search session. Besides these two metrics, we count the number of days that each consumer logs into the Taobao app during the same period.

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,

[36] 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).

2011; Pfeiffer et al., 2020). We further classify consumers into two groups based on their average time spent in a search session before the category refinement. Appendix Table A6 summarizes the average length of a consumer's search session in our data. We define consumers whose clicking time is above the median of the population as *exploratory searchers*, who tend to scan and browse in a search environment without predefined goals. The rest of the 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 a difference-in-differences model estimation similar to our main analysis, with the dependent variables being 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 conducted more searches, used more search queries, and logged into the platform more frequently in the week following their first view of the more-precise search results. The positive experiences of quickly getting what they want made them more likely to return to the platform for shopping. In contrast, we find a significant decrease in engagement with the platform among exploratory searchers when the search precision increased. Their average number of days visiting the platform during a week decreased by 4.1%; their number of searches in a week decreased by 5.5%; and their number of search queries decreased by 5.6%. These consumers mostly enjoy browsing and scanning various product categories on the platform, and too-precise search results may take away their enjoyment of exploring and cause them to spend less time on the platform.

[Insert Table 6 about here.]

7.2 Cross Selling

The search results' precision 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 top co-purchases based on treated consumers' transaction histories. We also calculate the total purchase amount across the leaf categories that share the same first-tier or second-tier category with the "Smart Trash Can." [37] 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 related to the category "Smart Trash Can" after the category refinement. Improving search precision can boost goal-directed searchers' confidence in 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 from exploratory searchers if the search results are too precise. Results in Table 7 indicate that these consumers were likely to decrease their spending on the relevant product categories by about 1-2% in the following week when search precision improved.

[Insert 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.

8.1 Mechanism: 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.

---

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

To test this hypothesis, we calculate the search traffic of the query "Smart Trash Can" distributed to the categories "Smart Trash Can" and the "Generic 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.[38] Figure 6 shows the ratios of weekly search traffic distributed to these two categories, respectively, for the query "Smart Trash Can" throughout 2019. Before the category refinement took place on April 18, 2019, less than 2% of search traffic was allocated to the category "Smart Trash Can."[39] After the category refinement, this number rose to over 15% and increased to nearly 25% at the end of 2019.

These results suggest that the search engine provided significantly more search traffic to the refined product category. Before the category refinement, consumers who searched for a niche product such as the "Smart Trash Can" were more likely to see only "Generic Trash Can" in the search results. With more-refined and better-targeted search results due to an improved search engine, consumers had a better chance of finding the right products.

[Insert Figure 6 about here.]

8.2 Supply-Side Response: Stronger Long-Tail Effect

The previous analysis focuses primarily on the reactions from the demand side, trying to understand how the improvement in search quality affects consumers' search outcomes, search intensity, and engagement with the platform. However, sellers can also re-optimize their competition strategies in the long run as a result of the changes in search algorithms. We now look at the supply side of the market and examine whether and how sellers adapt their behaviors after the category refinement.

We find that the weekly number of smart trash cans on the platform increased by 150% as search quality improved. The higher volume of search traffic attracted many more sellers to put their products into this niche category. We further extract all sellers who showed up in the search results for the query "Smart Trash Can" and divide them into two groups: sellers who listed products under the niche category "Smart Trash

[38] 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, Fan, and Tan, 2020). We get similar results when using these two measures of search traffic.

[39] Taobao 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 implement category refinement on April 18, 2019.

Can" and sellers whose products belonged to the category "Generic Trash Can." Figure 7 suggests that sellers with fewer ratings were 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. Therefore, entering a narrow category with limited search traffic may not generate enough profits to cover their monthly salary expenses.

[Insert Figure 7 about here.]

Overall, our findings suggest that niche product sellers, as a whole, will gain more profits as the quality of search improves. Both the total transaction volume and revenues from smart trash cans significantly increased after the category refinement. By reducing the search friction on the platform, the search engine made it easier for relevant sellers to reach consumers who may have a higher willingness to pay for their products. As a result, the total revenues of the refined product category nearly doubled in our study period. Sellers who sell niche products are significantly better off with the advance of the search technology. With an increasing number of specialized products that cater to more-specific segments of consumers, small and new sellers gain more market share, and the distribution of sales becomes less concentrated.[40]

9. Conclusion and Implications

This paper explores the role of the e-commerce platform in improving search precision and measures 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 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 over a longer time horizon. This negative search deterrence effect is more pronounced for exploratory searchers who have only a particular

[40] Our findings are consistent with the literature on how online search tools or recommendation systems affect sales distribution (Brynjolfsson, Hu, and Simester, 2011; Bar-Isaac, Caruana, and Gunat, 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, makes consumer search more targeted. The same rationale can be applied to our setting.

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 e-commerce search design that has not been documented in this literature.

Our findings have 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 to improve information retrieval system results (Dumais and Chen, 2000; Xing et al., 2008; Bilal, 2012; Farhoodi, Ghidary, and Yari, 2013). Our analysis shows, empirically, that improving search engine algorithms can, indeed, better fulfill consumers' needs and help generate revenues. Meanwhile, platforms may be well advised to consider the potential negative impact on consumer engagement, especially if they also derive revenues beyond sales (e.g., advertisements). In addition, our results indicate that e-commerce companies should fully leverage the textual data from online search queries to understand consumers' heterogeneous shopping intents. Product search engines are quite different from information search engines such as Google or Bing (Ghose, Ipeirotis, and Li, 2014). On e-commerce platforms, people do not 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 analyses of consumer heterogeneity based on the texts of their search queries (§6.3) illustrates the value of this approach.

Our study of platform search design has limitations that imply directions for future research. First, while we have illustrated the exploitation-exploration tradeoff, the optimal design of search precision and the optimal depth of the 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, e-commerce platforms develop many advanced information technologies besides search algorithms to guide consumers' product discovery process. It is crucial to understand how information, such as personal data, and other technologies, such as recommendation algorithms, AI chatbot, or live streaming, moderates the impacts of search algorithms. [41]Finally, though as complete and granular as possible, our search data are confined to one particular domain and, thus, fail to capture search activities across other e-commerce websites and offline retailers. Datasets that track individual consumers' entire search path, both online and offline, will further enrich our insights and 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.

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).

Angrist, J.D. and Pischke, J.S., 2008. *Mostly harmless econometrics: An empiricist's companion*. Princeton university press.

Autor, D.H., 2003. Outsourcing at will: The contribution of unjust dismissal doctrine to the growth of employment outsourcing. *Journal of Labor Economics*, *21*(1), pp.1-42.

Arazy, O. and Woo, C., 2007. Enhancing information retrieval through statistical natural language processign: A study of collocation indexing. *Mis Quarterly*, pp.525-546.

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.

Basu, A., Bhaskaran, S. and Mukherjee, R., 2019. An analysis of search and authentication strategies for online matching platforms. *Management Science*, *65*(5), pp.2412-2431.

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).

[41] For example, a new study by Sun et al. (2020) evaluates the importance of the use of personal data in Taobao's recommendations to consumers. For another example, Sun et al. (2019) examines the impact of Taobao's voice AI product on consumer search and purchase behaviors. These studies continue to further our understanding of search technologies.

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.
Broder, A., 2002, September. A taxonomy of web search. *SIGIR Forum*, 36(2), 3-10
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., Hu, Y. and Rahman, M.S., 2009. Battle of the retail channels: How product selection and geography drive cross-channel competition. *Management Science*, *55*(11), pp.1755-1765.
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.
Card, D. and Krueger, A.B., 1994. Minimum wages and employment: a case study of the fast-food industry in New Jersey and Pennsylvania. *American Economic Review*, 84(4), pp.772-793.
Chetty, R., Friedman, J.N., Leth-Petersen, S., Nielsen, T.H. and Olsen, T., 2014. Active vs. passive decisions and crowd-out in retirement savings accounts: Evidence from Denmark. *The Quarterly Journal of Economics*, *129*(3), pp.1141-1219
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.
Dewan, S. and Ren, F., 2007. Risk and return of information technology initiatives: Evidence from electronic commerce announcements. *Information Systems Research*, *18*(4), pp.370-394.
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.*
Du, X., Su, M., Zhang, X. and Zheng, X., 2017. Bidding for multiple keywords in sponsored search advertising: Keyword categories and match types. *Information Systems Research*, *28*(4), pp.711-722.
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., 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.
Jansen, B.J., Booth, D.L. and Spink, A., 2008. Determining the informational, navigational, and transactional intent of Web queries. *Information Processing & Management*, *44*(3), pp.1251-1266.

Johar, M., Mookerjee, V. and Sarkar, S., 2014. Selling vs. profiling: Optimizing the offer set in web-based personalization. *Information Systems Research*, *25*(2), pp.285-306.
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
Li, Z. and Agarwal, A., 2017. Platform integration and demand spillovers in complementary markets: Evidence from Facebook's integration of Instagram. *Management Science*, *63*(10), pp.3438-3458.
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*, 31(3), pp. 653-1036
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*.
Rose, D.E. and Levinson, D., 2004. Understanding user goals in web search. In *Proceedings of the 13th international conference on World Wide Web* (pp. 13-19).
Roshdi, A. and Roohparvar, A., 2015. Information retrieval techniques and applications. *International Journal of Computer Networks & Communications Security*.
Schumpeter, J. A. 1934. The theory of economic development. *Harvard: Economic Studies*, *46*, pp.1600-0404.
Shen, D., Ruvini, J.D. and Sarwar, B., 2012. Large-scale item categorization for e-commerce. In *Proceedings of the 21st ACM international conference on Information and knowledge management* (pp. 595-604)
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), 213-22.

Sun, C., Shi, Z.J., Liu, X., Ghose, A., Li, X. and Xiong, F., 2019. The Effect of Voice AI on Consumer Purchase and Search Behavior. *Working Paper*.
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.
Sun, T., Yuan, Z., Li, C., Zhang, K. and Xu, J., 2020. The Value of Personal Data in Internet Commerce: A High-Stake Field Experiment on Data Regulation Policy. *Working Paper*.
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.
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).
Wang, W., Li, B. and Luo, X., 2019. AI Agents for Sequential Promotions: Combining Deep Reinforcement Learning and Dynamic Field Experimentation. *Working Paper*.
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).
Xu, K., Chan, J., Ghose, A. and Han, S.P., 2017. Battle of the channels: The impact of tablets on digital commerce. *Management Science*, *63*(5), pp.1469-1492.
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.
Zhang, J., Adomavicius, G., Gupta, A. and Ketter, W., 2020. Consumption and Performance: Understanding Longitudinal Dynamics of Recommender Systems via an Agent-Based Simulation Framework. *Information Systems Research*, *31*(1), pp.76-101
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: Search 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: Columns (1)-(2) and (4)-(5) show means with standard deviations in parentheses. Columns (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) models and the triple differences models, 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 the number of listings a consumer views during a search session. The other two are the number of listings on which a consumer clicks and the total time she spends on these clicked listings. We take logs for all 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***	0.00035***	-0.0433*	-0.0371***	-0.0619***
	(0.0003)	(0.00006)	(0.0182)	(0.0083)	(0.0179)
After × Smart Trash Can × Specific Needs	0.0316***	0.00041	-0.0528	-0.0437	-0.1063
	(0.0015)	(0.00028)	(0.0354)	(0.0321)	(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 Tables 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 includes only 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 consumers 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**	0.0367**	0.0157***
	0.0124	0.0138	0.0026
After × Smart Trash Can × Exploratory Searchers	-0.0547*	-0.0559*	-0.0414***
	0.0263	0.0263	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 week after seeing the more precise search results. We take logs for all 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 week after seeing the more precise search results. We take logs for all 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.

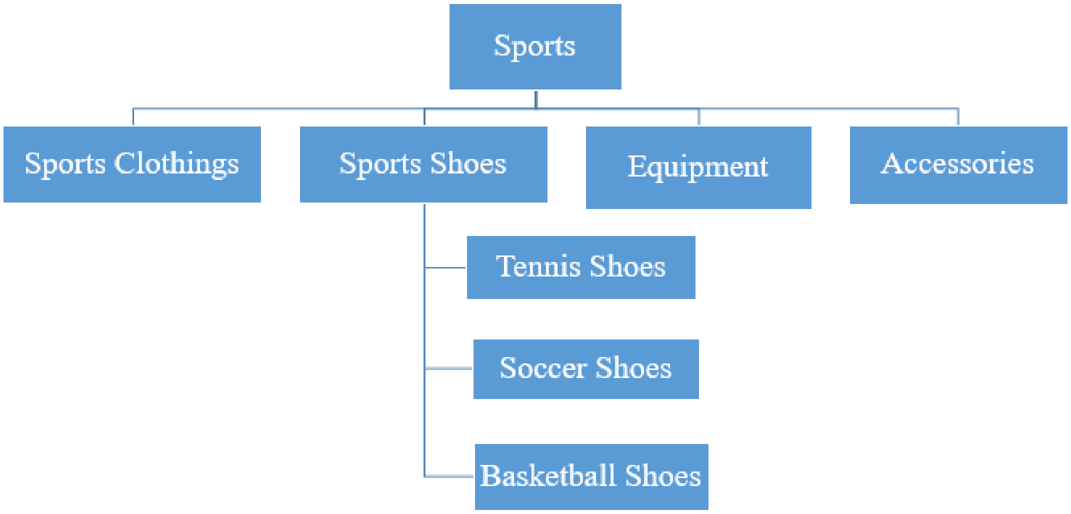


Figure 1: An Example of Category Hierarchy on E-commerce Platforms

Figure 2(a): Category Selection on Alibaba

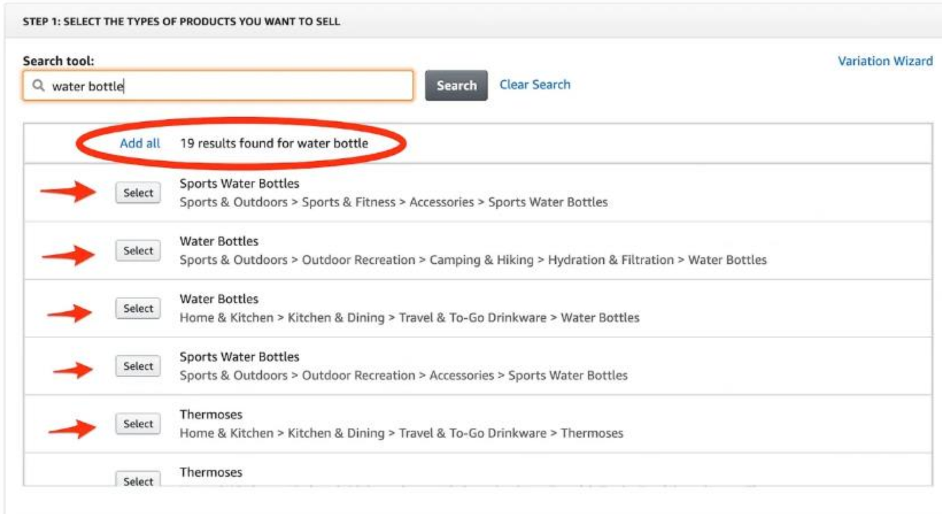
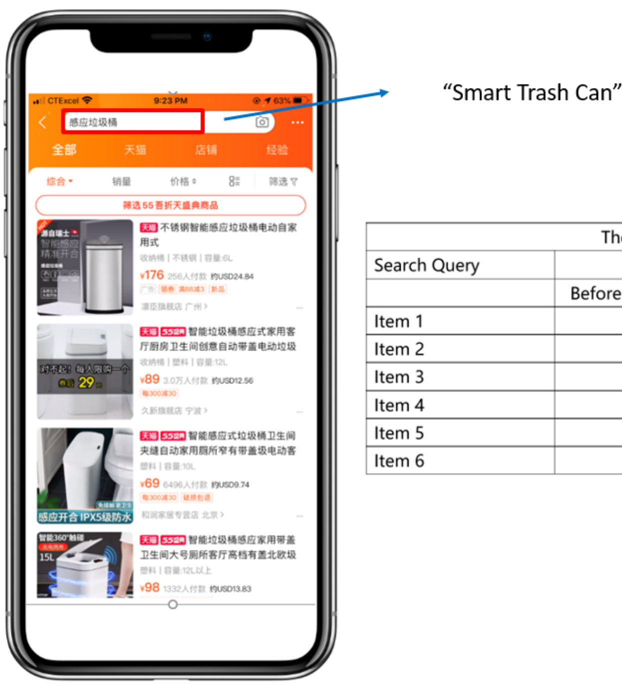


Figure 2(b): Category Selection on Amazon

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



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 table on the right to be what consumers actually see on their mobile phones before and after the category refinement, respectively.

Figure 4: A Typical Online Shopping Journey on Alibaba

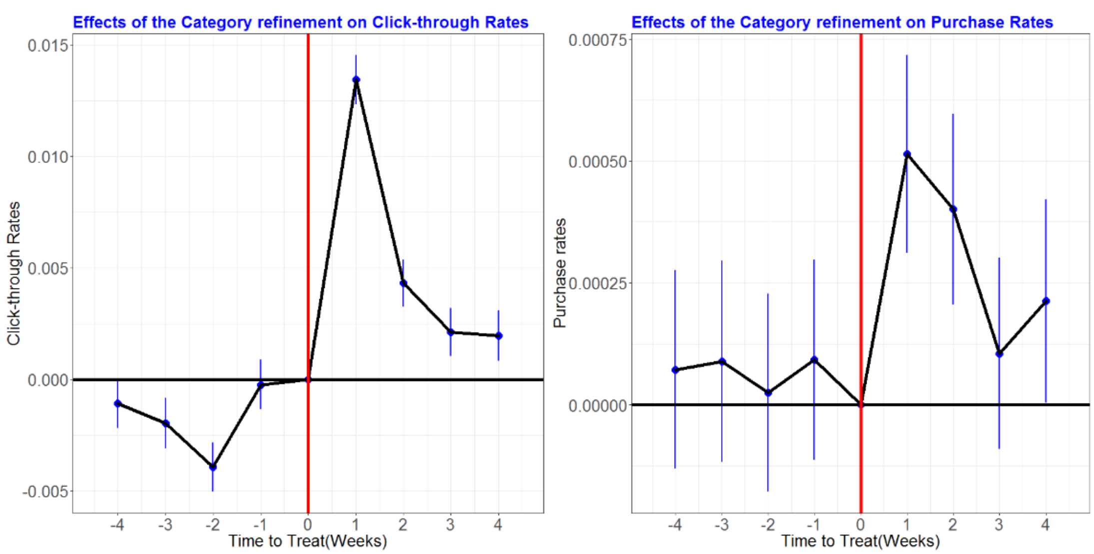


Figure 5 An Event Study of the Effect of Category Refinement on Matching Outcomes

Note: Figure 5 shows DID coefficients and 95-percent confidence intervals from the 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.

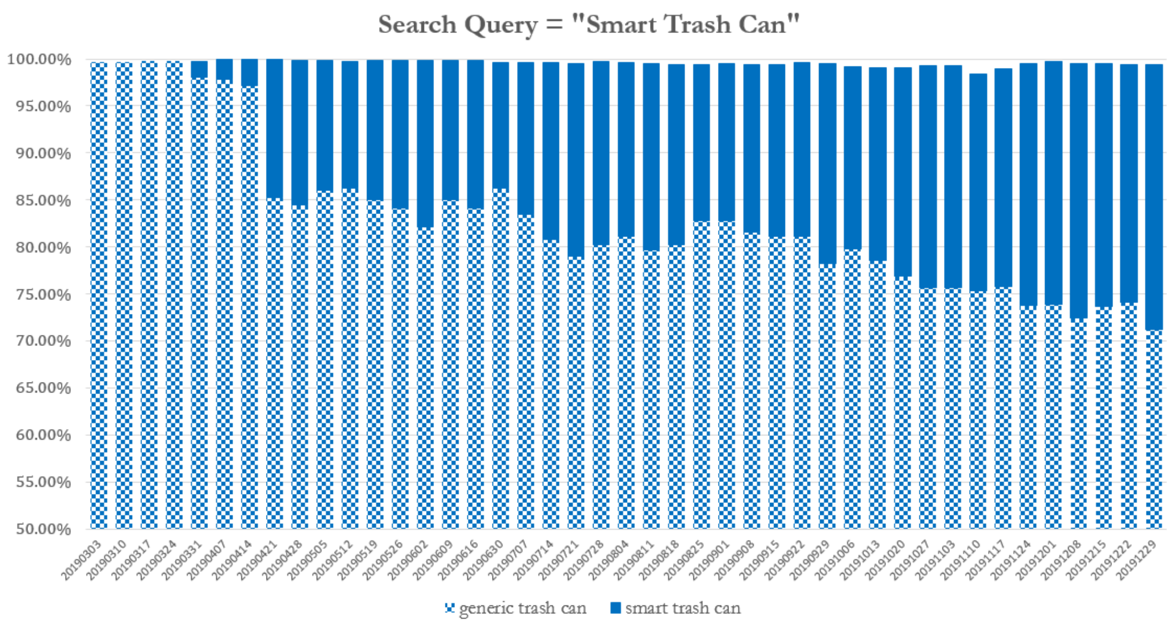


Figure 6: An Example of Search Traffic Allocation

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 on 04/18/2019. The half-filled bars represent the ratio of search traffic allocated to the products listed under the category "Generic Trash Can" as a whole. The fully filled bars represent the ratio of search traffic allocated to the products belonging to the category "Smart Trash Can" as a whole.

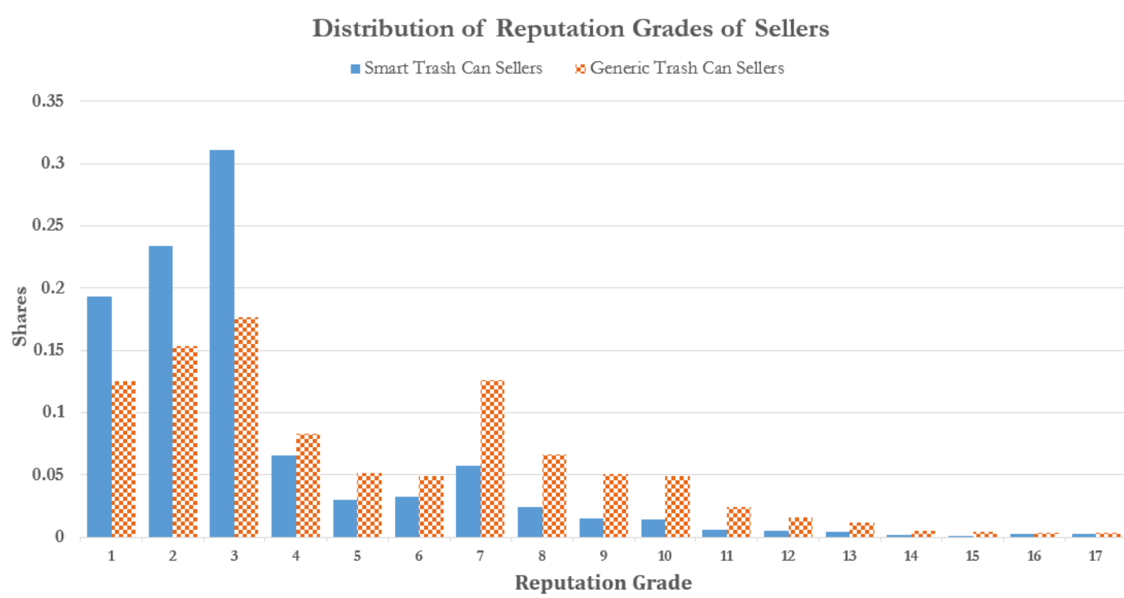


Figure 7: Compare the Distribution of Reputation Grades by Groups

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 "Generic 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
02/21/2019	E-cigarette/E-liquid	E-cigarette accessories E-liquid E-cigarette
	Candy	Wedding Candy Other Candies
03/18/2019	Other Vegetables	Red pepper Broccoli Other fruits
04/17/2019	Tissue	Paper towel Napkin Tissue
	Trash Can	**Smart Trash Can** **Generic Trash Can**
05/17/2019	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 the platform in the last week	5.85	5.86	0.01
	(1.60)	(1.58)	0.15
# Days the consumer logged into the platform in the last month	23.80	23.72	0.08
	(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 record 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 the 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: Four-month 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 before the Category Refinement

	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

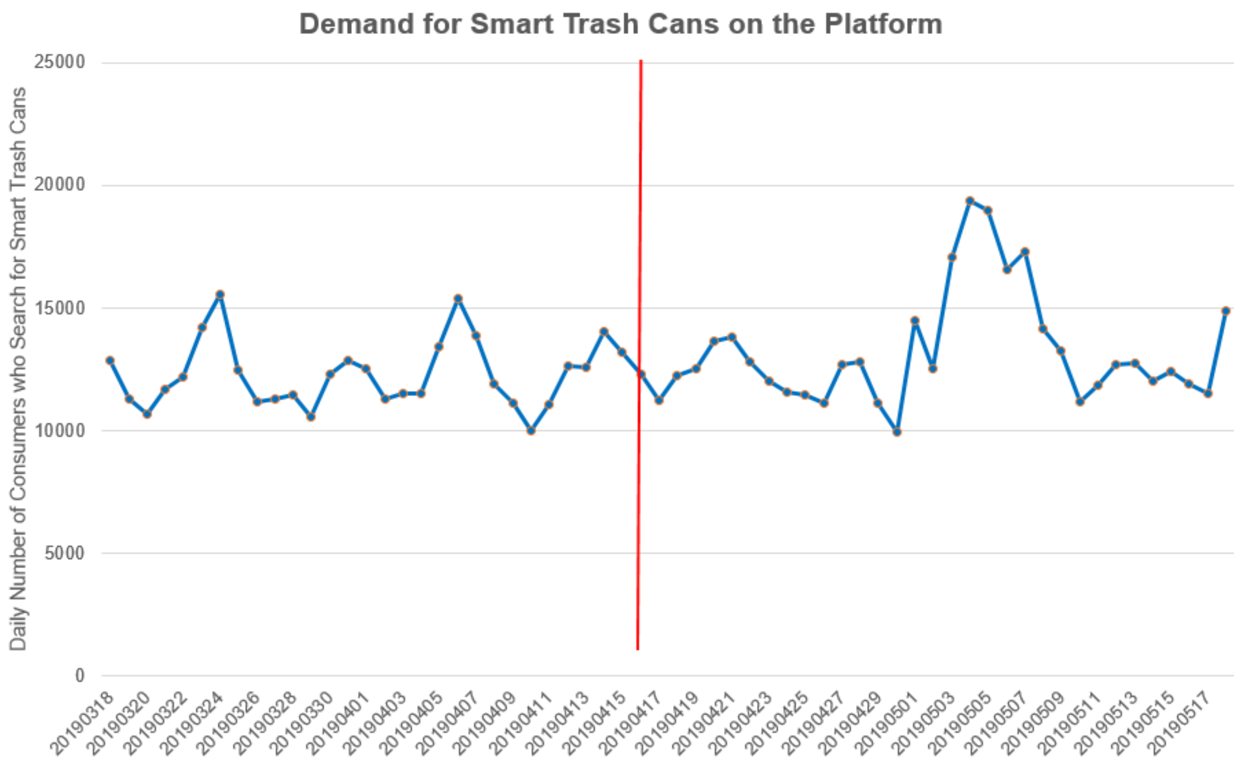


Figure A1: Demand for Smart Trash Cans on the Platform throughout the Sample Period

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.