NBER WORKING PAPER SERIES

# PLATFORM INFORMATION PROVISION AND CONSUMER SEARCH: A FIELD EXPERIMENT

Lu Fang
Yanyou Chen
Chiara Farronato
Zhe Yuan
Yitong Wang

Working Paper 32099
http://www.nber.org/papers/w32099

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
February 2024

Zhe Yuan would like to acknowledge financial support from the NSFC (Grants 72203202, 72192803 and 72141305). Lu Fang would like to acknowledge financial support from the NSFC (Grant 72192803). Yitong Wang is an employee at the company that shared the data for this research. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Platform Information Provision and Consumer Search: A Field Experiment
Lu Fang, Yanyou Chen, Chiara Farronato, Zhe Yuan, and Yitong Wang
NBER Working Paper No. 32099
February 2024
JEL No. D81,D83,L2,L81,L86

## ABSTRACT

Despite substantial efforts to help consumers search in more intuitive ways, text search remains the predominant tool for product discovery online. In this paper, we explore the effects of visual and textual cues for search refinement on consumer search and purchasing behavior. We collaborate with one of the largest e-commerce platforms in China and study its roll out of a new search tool. When a customer searches for a general term (e.g., "headphones"), the tool suggests refined queries (e.g., "bluetooth headphones" or "noise-canceling headphones") with the help of images and texts. The search tool was rolled out with a long-run experiment, which allows us to measure its short-run and long-run effects. We find that, although there was no immediate effect on orders or total expenditures, the search tool changed customers' search and purchasing behavior in the long-run. Customers with access to the new tool eventually increased orders and expenditures compared to those in the control group, especially for non top-selling products. The purchase increase comes from more effective searches, rather than an increase in activity on the platform. We also find that the effect is not only driven by the direct value of suggested searches, but also by customers indirectly learning to perform more effective searches on their own.

Lu Fang
The Rural Development Academy
866 Yuhangtang Rd.
West Lack District
Hangzhou, Zhej 310058
P.R.China
fl_fanglu@zju.edu.cn

Yanyou Chen
University of Toronto
Max Gluskin House
150 St. George Street
Toronto, ON M5S 3G7
Canada
yanyou.chen@utoronto.ca

Chiara Farronato
Harvard Business School Morgan
Hall 427
Soldiers Field
Boston, MA 02163
and NBER
cfarronato@hbs.edu

Zhe Yuan
Zhejiang University
yyyuanzhe@zju.edu.cn

Yitong Wang
yitong_wang2019@outlook.com

# 1 Introduction

Online consumer search is dominated by text queries. To make search easier, platforms such as Amazon and Google have added autocomplete functionalities to their search bars. As consumers start typing, they are offered a variety of suggestions, either from their own past search behavior or from popular searches performed by other customers.[1] In addition to text suggestions, Google launched Lens in 2017, with the hope that it would facilitate visual searches that are otherwise difficult to describe in words.[2]

Although technology companies have put efforts to help consumers search in more intuitive ways, there is limited empirical evidence as to the value of these efforts. To shed light on the value of search refinement tools for consumers online, we exploit the launch of a search tool recommending more refined searches through text and pictures on one of the largest e-commerce platforms in the world. The roll out had two features that are helpful for our analysis. First, the new search tool was randomly made available to a subset of the platform's consumers, allowing us to estimate the causal effects of the tool. Second, the experimental period lasted for approximately ten months, allowing us to quantify both the short-run and long-run effects of the search tool.

We find that the search tool was immediately effective at changing consumers' search behavior. On the first day they entered the experiment, 5.5% of customers in the treatment group searched for queries that were suggested by the search tool, compared to only 1% of customers in the control group. Despite the change in search behavior, the tool had no immediate effect on consumer transactions, measured as either the number of orders placed or total expenditures.

The long-run effects paint a very different picture of the effect of the search tool. In the following 24 weeks since entering the experiment, customers in the control group spent 3.2% more and completed 1.6% more orders compared to the control group. These results are not explained by increased activity on the platform (e.g., more searches or product views), implying that search became more efficient after the introduction of the search tool. We find evidence that the increase in consumer purchases does not only come from searches directly affected by the new search tool, but also spills over to other searches on the platform. We confirm that at least part of the spillover effects come from consumers learning to perform more specific searches. Our findings reveal a notable increase in customer satisfaction, as evidenced by higher positive ratings from customers and a reduced rate of product returns.

Our paper highlights the important role that search design plays in helping consumers

[1] See https://support.google.com/websearch/answer/7368877?hl=en as an example of how autocomplete works on Google Search.

[2] https://lens.google/

identify what they want and find it. Our partner platform, like many other e-commerce sites, allows consumers to input their search queries in text form. The platform then returns a list of products matching the query, ranked according to proprietary algorithms. When conducting searches, consumers typically have two types of tools to refine their search. First, they can use pre-defined filters (e.g., brand or price filters) to exclude certain products. Second, consumers can sort products according to criteria other than the default ranking, for example recency – where more recently added products are displayed first – or price – where cheaper products are displayed first. These search functionalities operate most effectively when customers have clear preferences, i.e., they know what they are looking for and how to clearly describe it.

Two potential challenges arise in a text-based search process. First, customers may have a good understanding of their needs but lack knowledge of the corresponding search terms (Liu and Toubia, 2020). For example, a user may know they want cordless headphones, but may not know that *bluetooth* is the typical technology to connect such headphones to their electronic devices. This challenge, known as *demand expression*, arises when customers struggle to articulate their requirements effectively while conducting searches. Demand expression seems to be an important challenge in online search, at least judging from the number of websites with tips for more efficient searching strategies (Markey, 2019).[3] Yet, the existing literature remains limited (Lazonder, 2005).

When consumers find it difficult to describe in words what they want, it may be beneficial for platforms to incorporate intuitive and user-friendly interfaces, along with appropriate visual cues and descriptive text, in order to bridge the gap between customer needs and effective search queries. By facilitating better demand expression, platforms can enhance the overall customer experience and ensure customers find the products or information they want.

Second, customers might have a general idea of what they want, but lack specific information about the characteristics of the products available, and hence of the products they ultimately want. For example, a user may know they want headphones, but don't know that they can choose between over-ear or in-ear headphones. This challenge is often referred to as *demand formation*. Prior research has demonstrated that recommendation systems can influence consumers' consideration sets, help them identify what they want (Häubl and Murray, 2003; Fong, 2017; Wan et al., 2023; Yuan et al., 2023) and how much they are willing to pay (Adomavicius et al., 2013, 2018, 2019). Such results support the hypothesis

[3]See, for example, https://www.lifewire.com/web-search-tricks-to-know-4046148, https://www.techrepublic.com/article/10-tips-for-smarter-more-efficient-internet-searching/, https://www.indeed.com/career-advice/career-development/internet-search-tips, or https://mediasmarts.ca/tipsheet/how-search-internet-effectively.

that customers may develop their demand while searching, rather than searching for what they already know they want.

To address these issues, some platforms have adopted auto-complete technology, whereby consumers starting to type a query may be presented with relevant query suggestions, from their individual past history or from aggregate search behavior. Our collaboration allows us to go one step further and explore how incorporating guidance as a combination of pictures and words can help address both demand expression and formation challenges. As we emphasize in the next section, there is remarkably limited empirical work identifying the effects of improved search functionalities, with the exception of Lei et al. (2023) and Tong et al. (2023). Even less evidence exists on the long-term effects of such functionalities and the learning value they provide to consumers, who can learn to improve the effectiveness of their searches even when those search functionalities are not available.

Our results have important implications for the design of online search mechanisms. On one hand, consumers have private information over what they want to search online. On the other hand, they can benefit from recommendations that help refine their searches and inspire their interests towards products they may not ex-ante know they want. Our results suggest that the current design of search mechanisms may still overly rely on consumers' prompts, despite platforms having extensive knowledge about consumer preferences, in the aggregate as well as at the individual level. We expand on these topics in the concluding section.

The paper is structured as follows. Section 2 describes the existing literature to which our paper contributes. Section 3 presents the institutional setting, the experiment, and the data available. Section 4 focuses on our empirical approach and results, which are divided into short-run and long-run results. Finally, Section 5 concludes the paper, highlighting the managerial implications of our results.

# 2 Literature

Our paper contributes to the literature on customer search behavior and search costs, on platform design, and on field experiments. Since at least Gardner (1970), Weitzman (1979), and Rothschild (1974), researchers have been interested in understanding how people search. The advent of search engines and digital platforms have allowed empirical tests of the theories (Santos et al., 2012; Ursu, 2018), as well as quantifications of search frictions (Ellison and Ellison, 2009; Lee and Musolff, 2021). More recently, Bronnenberg et al. (2016) describe online consumers' search behavior while Seiler (2013) shows that search frictions significantly impact purchasing behavior. Choi et al. (2018) focus on the unexpected consequences

of lowering search frictions. In this paper, we identify consumers' inability to express in words what they are looking for as a source of frictions, and how visual and textual cues, designed to help consumers refine their searches, can be an effective solution. Although we cannot separate the role of visual and textual refinements, existing work has demonstrated that pictures play an important role in facilitating consumers' information acquisition and processing (Blanco et al., 2010; Wu et al., 2021).

There is recent growing research on platform design, specifically on how platforms present relevant options and what type of information they disclose about them. In the context of eBay, Dinerstein et al. (2018) is one of the earliest works that looks at how search ranking algorithms play a critical role in reducing search frictions and changing competition, ultimately determining market outcomes and welfare. More recent studies on the effects of changing how products are presented to consumers include Chen et al., 2023 and Yang et al., 2023. Fradkin (2017) finds large consumer benefits arising from room availability tracking and filtering on Airbnb. Filippas et al. (2023) find similar benefits from improving information disclosure about professionals' availability to take more jobs. Chen and Yao (2019) use click-stream data on hotel bookings to find sizable consumer benefits from search refinement tools, such as sorting and filtering. A crucial assumption underlies the ample work estimating the effects of disclosing information about products and services and the effects of changing the order of search results: consumers are assumed to know what they want and how to describe it. The behavioral literature has however found limitations to this assumption (e.g., Kamenica et al., 2011). Our paper confirms that consumers sometimes find it difficult to describe in words what they want.

There is more limited work on estimating the effects of tools that help consumers better discover and express their preferences. Lei et al. (2023) is one of the few papers quantifying the positive effects of auto-complete on consumer search. They leverage an experiment with a small search engine platform, which removes access to search recommendations from the API of a larger competitor. The authors find large benefits of the API in helping consumers find what they want. Similarly, Tong et al. (2023) leverage an experiment on a food delivery platform to show that query recommender systems increase the probability that customers place a food order, at least in the short-run. Häubl and Trifts (2000) conduct a controlled experiment using a simulated online store to show that interactive tools designed to assist consumer search have strong positive effects on purchase decisions. Our study adds to this body of work by revealing the impact of offering a search refinement tool on consumer search and purchasing decisions, not just in the short-run, but over an extended period of time.

The majority of the research on the value of search recommendations, such as Sun et al. (2023) and Chiou and Tucker (2017), focuses on the role of consumer data to help offer

personalized results. But consumer data can help further refine searches, by for example, identifying new search filters or search tools (Jiang and Zou, 2020). Our paper contributes to this latter line of research by highlighting the role of visual and textual suggestions in guiding consumer search. Our ability to observe the entire search and purchasing funnel allows us to shed light on the mechanisms through which search refinement tools benefit consumers, by increasing the likelihood that consumers find what they want for any individual search, and by teaching consumers to more effectively search on their own even when those tools are not available.

Our results that tail and niche products gain more from the introduction of search refinement tools relates to the extensive exiting literature exploring how the Internet reshapes market structure and concentration, particularly when comparing sales of popular products versus niche products (Elberse and Oberholzer-Gee, 2006; Bar-Isaac et al., 2012). Fleder and Hosanagar (2009) employ a theoretical model to investigate the effect of recommendation systems on sales diversity, and predict that recommendation systems based on sales and ratings tend to decrease sales diversity and promote product concentration. Although that may be true of baseline recommender systems, our findings suggest that search tools like the one we study may actually provide a correcting mechanism against sales concentration. Our findings align more closely with work by Brynjolfsson et al. (2011). They demonstrate that e-commerce and online search technologies allow consumers to discover products that better match their preferences, which in turn imply that niche products (i.e., the long tail) can capture larger market share than in the offline world. Notably, this long tail effect is not solely attributed to the expansion of product variety but also partially caused by the lower search costs online.

Finally, our paper also highlights the importance of running long-term experiments to identify the equilibrium effects of product changes (Gupta et al., 2019). In doing that, we relate to the literature on long-term experiments (Goli et al., 2021; Huang et al., 2018) and approaches to infer long-term outcomes from short-term proxies (Athey et al., 2019). We find that short-term results may be very different from long-term results. The typical risk of short-term experiments is that one may find positive short-term effects, but null or negative effects in the long-run (Kohavi et al., 2012). Our specific case highlights the opposite risk, i.e., improvements in platform design that take time to emerge.

# 3 Data and Institutional Details

We collaborate with one of the largest e-commerce platforms in the world, which we keep anonymous as part of our research agreement. Given the large variety of products available,

search tools on e-commerce platforms like our partner play a crucial role in helping customers find products that match their needs. In this section, we describe the search tool that our collaborating platform created, how they experimentally launched it, and the data we have available to study its effects.

## 3.1 The Picture-Text Search Tool and Its Experimental Roll-Out

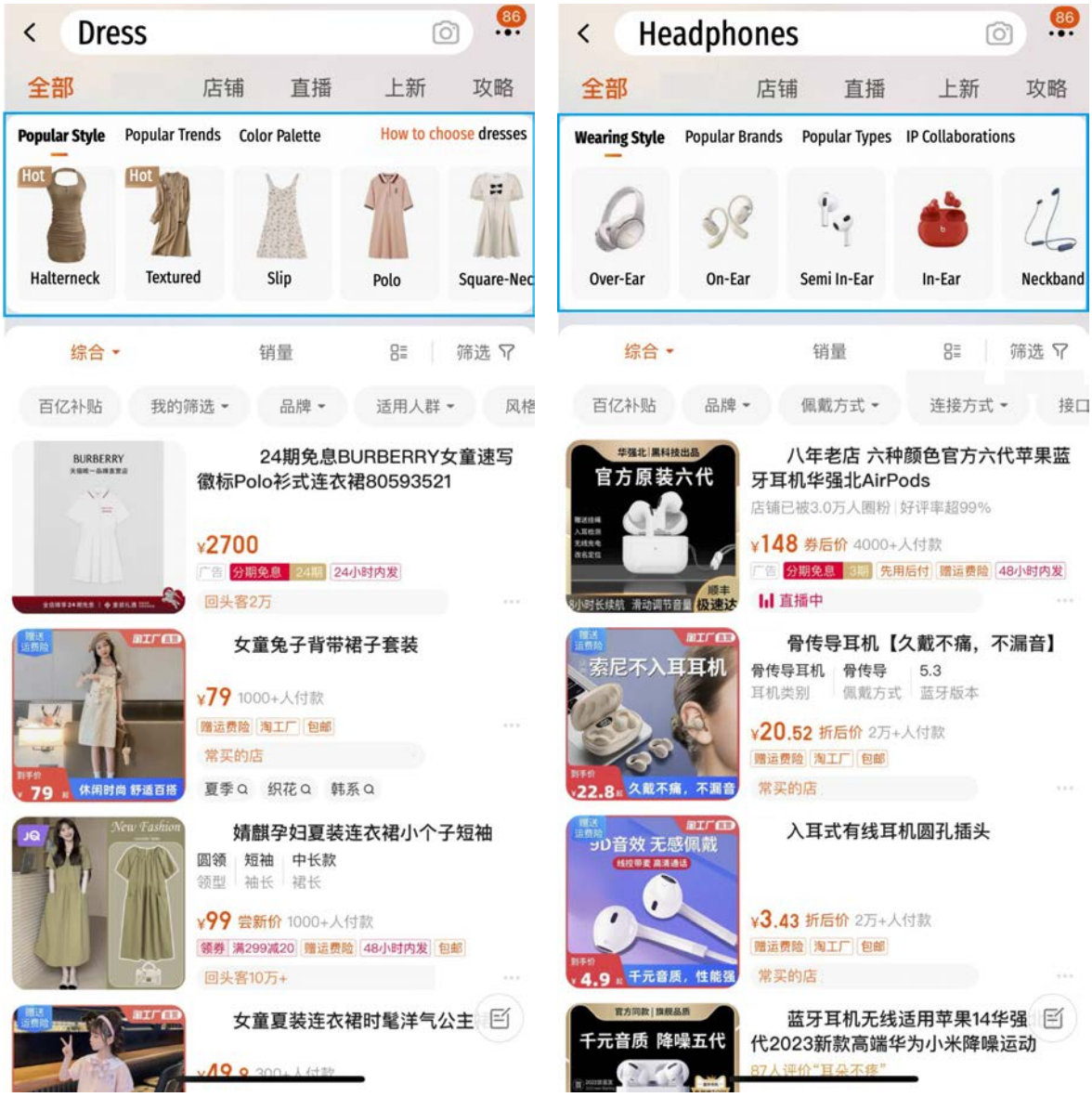
The collaborating platform has millions of sellers and hundreds of millions of customers active on any given month, and billions of products listed on any given day. Customer search plays a crucial role on this platform. Over 30% of purchases can be linked to a search that immediately precedes that purchase. This share is likely an underestimate of the role of search for purchases given that many consumers may add products to cart and then purchase them later.

The focus of our study is a new search tool that suggests a combination of picture and textual recommendations for the consumer to refine their searches. We call this tool *Picture-Text Guidance* (PTG henceforth) in the rest of the paper. Figure 1 illustrates how PTG works. When a customer enters a query that is a candidate for PTG, such as "Dress" on the left panel of Figure 1, the platform's search engine presents the consumer with two levels of sub-categorization of products related to the general search term. The first level presents broad dimensions for classifying relevant products. In the dress example, the picture shows "Popular Style," "Popular Trends," and "Color Palette." The second grouping level is presented as a series of pictures with the corresponding descriptive words. In the figure, the pictures correspond to dresses grouped by "Popular Style": Halterneck, Textured, Slip, Polo, and Square-Neck. The right panel of Figure 1 provides an analogous example for headphones.

Customers can click on any of the PTG elements to refine their search. When they click on one of those elements, the search engine will automatically refine the search query to reflect the finer subset of relevant products. For example, if the customer clicks on the picture for the halterneck dress, the word "Halterneck" is added to the search box at the top. Instead of returning results matching the query "Dress," the engine will thus return results matching the query "Dress Halterneck."[4]

Although we cannot disclose the details of the proprietary algorithms, the set of queries that are candidates for this refinement tool were identified by the product team based on the popularity of consumer searches, and the possibility to break down those searches into narrower queries. The candidate queries for PTG are selected based on their popularity,

[4]Appendix Figure A.1 presents more details about PTG.




(a) Dress (b) Headphones

Figure 1: Illustration of the Picture-Text Guidance (PTG) Search Tool.

determined by factors such as the number of consumers who have searched for a particular query. For example, "Dress" is among the top 1% queries in terms of cumulative searches conducted by consumers in 2021. Queries with higher popularity tend to represent a consumer's initial idea or a general expression of their needs. Therefore, identifying refined recommendations to these popular search queries has the potential to assist consumers in expressing and forming their demands more effectively. The candidates for these finer and more specific queries are identified by a combination of a collaborative filtering model and human curation. For instance, "Halterneck" is chosen as a suggested query associated with "Dress" because it is often used by consumers in conjunction with "Dress."

Because of the substantial effort in identifying finer categories for the many search queries that customers search on the platform, and because some searches cannot be further broken down into subcategories, not all search queries are candidates for PTG. We thus categorize search queries into three types:

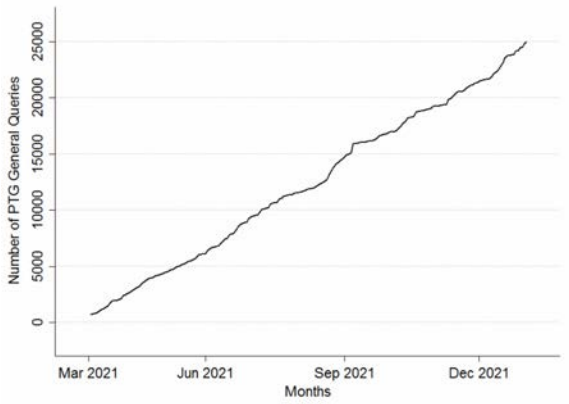
- *PTG general* queries refer to search queries that have been augmented with PTG. Examples of PTG general queries include "Dress" and "Headphones" as in Figure 1.

- *PTG specific* queries refer to search queries generated when a consumer clicks on the pictures following a search for a PTG general query. Examples of PTG specific queries include "Dress Halterneck" and "Dress Textured" on the left panel of Figure 1, and "Headphones Over-Ear" and "Headphones In-Ear" on the right panel. Note that customers can search for a PTG specific query by directly typing the words in the search box, not just by clicking on the picture provided by PTG. Our data will not be able to distinguish whether customers click on the PTG picture or type the query on their own.

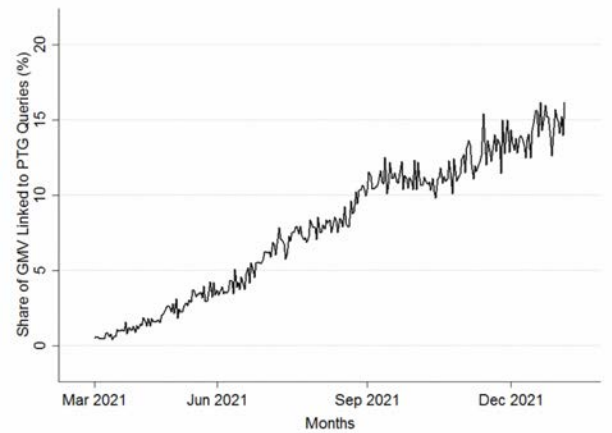
- *Non-PTG* queries refer to search queries that do not qualify for the PTG feature, such as "Squash Racket."

The platform launched PTG only on the mobile app in March 2021,[5] with a small number of PTG general and specific queries. Over the course of the following months, it progressively increased both the number of search terms classified as PTG general queries and the number of search terms classified as PTG specific queries. The left panel of Figure 2 shows that by December 2021, around 25,000 queries were classified as PTG general queries. The right panel of Figure 2 shows that PTG queries went from representing 0 to 15% of the gross merchandise volume (GMV henceforth) directly associated to a search query.

[5]Over 95% of consumers use the platform on mobile.




(a) Number of Queries Classified as PTG General Queries.




(b) Share of GMV Linked to PTG Queries (out of total GMV directly linked to searches).

Figure 2: Expansion of PTG Between March and December 2021.

During the roll-out of PTG between March and December 2021, the platform conducted a randomized field experiment to measure the effectiveness of the new search tool. All platform's customers (existing and new) were ex-ante randomly allocated to a control and a treatment groups with equal probability.[6] Upon entering a PTG general query, treatment and control customers saw different displays. Treatment customers saw the PTG search tool (picture and text suggestions in the blue rectangle in Figure 1) and could click on any of the search recommendations to refine their searches.[7] The control group did not have access to the PTG tool, and hence would not see the blue rectangle from Figure 1.

It is worth making two remarks. First, when customers searched for non-PTG queries, they would face the same standard search experience without the blue rectangle in Figure 1, regardless of whether they were in the treatment or control group. Second, because not all customers searched for PTG general queries, in our analysis we only include customers who searched for PTG general queries during the experimental period. The ex-ante random allocation ensures that focusing on this subset of users does not undermine our causal analyses.

The experiment lasted for ten months, from mid March until end of December 2021.

[6]Consumers who were not logged in when searching for products are not included in our experiment. This affects a small number of searches on mobile. Similarly, because the tool was only implemented on the mobile app, consumers who searched on the Web were not included in the experiment. Less than 5% of consumers use the platform on a web browser.

[7]Additional information regarding the PTG search tool and the potential responses of treatment group consumers to PTG are provided in Appendix A.

Since there is variation in the timing when customers first search for PTG general queries, we say that a customer *enters the experiment* on the first day during the experimental period when they search for a PTG general query.

This experiment proves very valuable for our goal of understanding whether and how search guidance tools help consumers better identify and describe what they want. The long experiment duration was driven by the fact that the search tool was progressively increasing its reach as new queries were included in PTG, but it provides a unique opportunity for us to measure both short-term and long-term effects of the new search tool, and evaluate the validity of the conclusions that would have been drawn if we had had access only to a short-run experiment.

## 3.2 Data

We obtain proprietary data from the platform. Although the roll-out of PTG continued past the end of 2021, we have access to data between mid March and Dec 31, 2021 (the *experimental period*). We restrict attention to treatment and control customers who reside in China and who performed a PTG general query during the experimental period.

Data are aggregated at the search level. For each search performed by a customer included in the experiment, information on the search terms allows us to classify the query into PTG general, PTG specific, or non-PTG. For each of the searches, we also have information on the following outcomes of interest: the number of products viewed in the search results (*views* henceforth);[8] the number of clicks on products returned in the search results (*clicks*); the number of purchases that were directly linked to the search (*orders*); and the total expenditures for those purchases (*GMV*, for gross transaction volume).

We augment the search-level data with product- and seller-level information. Specifically, for each of the products viewed, we obtain the product category and its seller identifier, in order to calculate sales rankings for both products and sellers. This additional information allows us to distinguish between more and less popular products or sellers, and how PTG affects consumer choices for different product and seller groups.

Similarly, we also augment the search-level data with customer-level information to compare customers in the treatment and control groups. Our sample includes 505,485 customers, half in the treatment group and half in the control groups. Table 1 confirms that the randomization was effective at allocating comparable customers into the two groups. On average,

[8]The number of product views is both a function of product availability for the specific search query, and of how much the customer continues to scroll past the initial results. Products are grouped into sets of about a dozen (we cannot disclose the exact number) – the first 12 results, the next 12 results, and so on. When a customer scrolls past a multiple of 12, an additional 12 products are added to the list of product views, as long as there are relevant products that remain to display.

Table 1: Covariate Balance

| | | Control Group $n = 252,737$ | Treatment Group $n = 252,748$ | P-value $(C = T)$ |
|---|---|---|---|---|
| Age Tier | Mean | 2.4853 | 2.4892 | 0.2654 |
| | Std Err | 0.0025 | 0.0025 | |
| City Tier | Mean | 3.8480 | 3.8518 | 0.4463 |
| | Std Err | 0.0036 | 0.0036 | |
| Number of Registered Years | Mean | 6.2573 | 6.2636 | 0.5044 |
| | Std Err | 0.0067 | 0.0066 | |
| Female | Mean | 0.5455 | 0.5454 | 0.9707 |
| | Std Err | 0.0010 | 0.0010 | |
| View in the Past 8 Weeks | Mean | 2585.4 | 2563.9 | 0.0865 |
| | Std Err | 8.95 | 8.80 | |
| Clicks in the Past 8 Weeks | Mean | 112.0 | 111.5 | 0.3835 |
| | Std Err | 0.39 | 0.39 | |
| Orders in the Past 8 Weeks | Mean | 4.53 | 4.55 | 0.4997 |
| | Std Err | 0.02 | 0.02 | |
| GMV in the Past 8 Weeks | Mean | 410.6 | 415.4 | 0.4888 |
| | Std Err | 5.73 | 4.06 | |

*Notes:* The table displays characteristics of customers included in the experiment. Customer characteristics refer to demographics – age grouping (where age is grouped in 10-year groupings, and 1 is assigned to the youngest 10-year grouping between 15 and 25 years old), city tier (where 1 is assigned to the largest cities in China, such as Beijing and Shanghai, 2 is assigned to cities like Hangzhou and Nanjing, all the way to tier 6, which includes the smallest towns and villages), tenure on the platform in years, the proportion of women – and behavior on the platform in the 8 weeks preceding their entry into the experiment – product views, clicks, orders, and GMV.

customers are between 25 and 35 years of age (denoted as age tier 2), they reside in large cities (3 denotes the third largest city-tier in China, out of a total of 6 tiers), they have been users of the platform for 6.3 years, and are 55% women. When it comes to customer behavior on the platform, Table 1 shows that in the 8 weeks preceding their entry into the experiment, customers viewed about 2,500 products, clicked on 112 products, purchased 4.5 of them, and spent CNY410-415 (almost $60).[9]

The next section describes our analyses, divided into a short-run and a long-run analysis. For the short-run analysis, we consider all the experimental customers in Table 1. For the long-run analysis, we restrict attention to customers entering the experiment between mid March and mid July 2021, allowing us to track them for almost 6 months until the end of 2021.

# 4 Empirical Approach and Results

We evaluate the effect of the PTG search tool on customer search behavior and purchase decisions. To do so, we conduct our analysis at the individual customer level (i.e., the randomization level) and estimate regressions of this form:

$$y_i = \beta \times Treat_i + \alpha_{c(i)} + \epsilon_i, \tag{1}$$

where $i$ denotes a customer in the experiment. Since customers enter the experiment when they first type a PTG general query, we control for their day of entry with cohort $c(i)$ fixed effects. $Treat_i$ is an indicator for whether the customer belongs to the treatment group, so the coefficient $\beta$ measures the causal effect of giving customers access to the search tool.

We estimate the regression for several outcomes $y_i$ tracking the customer behavior from search to purchase. We focus on the number of products each customer views, the number of clicks they make to those products, the total number of products purchased, and the overall expenditures linked to those purchases (GMV). These metrics are computed at the individual consumer level and aggregated over a designated time period, depending on whether we focus on the short-run or the long-run. To identify the mechanisms through which the effects materialize, we explore additional outcomes in the results section as needed.

We are interested in estimating the immediate effects of the search tool as well as the longer-term effects, which may include learning to perform more effective searches independently. In the short-run, we aggregate the outcomes of interest over the course of the first

[9]Note that the statistics in Table 1 do not necessarily reflect the usage characteristics of the entire population of platform customers, given that customers in our dataset are selected by the fact that they perform a PTG general query during the experimental period.

day when a customer enters the experiment. This allows us to use all customers who joined the experiment between mid March and end of December 2021 (505,485 customers). In the long-run, we aggregate the outcomes of interest over the course of 24 weeks following a customer entry in the experiment, which requires us to constrain the analysis to customers who entered the experiment between mid March and mid July 2021 (346,110 customers).[10]

Given recent concerns around using log transformations when outcomes can take the value zero (Chen and Roth, 2023), we estimate regressions in levels, and present their short-run and long-run estimates in the next two sub-sections.

## 4.1 Short-Run Results

This section focuses on outcomes measured on the day a customer enters the experiment. First, we show that the PTG search tool had a large and immediate impact on customers' search behavior. To do this, we estimate the effect on three separate outcomes: total number of searches conducted on the customer's first day in the experiment,[11] number of PTG general searches, and number of PTG specific searches.

We run regressions of Equation 1, and Table 2 displays the results. Column 1 shows that the search tool leads customers to perform 0.049 more queries compared to the baseline, which amounts to a 1.04% increase. This increase in searches comes solely from a rise in the number of PTG specific queries (column 3), which increase by 0.053, an almost identical coefficient to the estimate in column 1. Although this seems like a small increase in levels, the search tool effectively grows the propensity of consumers to perform PTG specific queries 5-fold. At least in the short-run, these additional PTG specific searches do not cannibalize PTG general (column 2) or non-PTG searches, which remain fairly constant.

Table 2 confirms that customers actively utilize the new search tool to perform narrower searches than in the absence of PTG, so our next step is to evaluate whether this change in search behavior translates into changes downstream, all the way to purchases. We thus estimate regressions as in Equation 1 for views, clicks, purchases, and expenditures on the customer's first day in the experiment.

Table 3 presents the results. None of the coefficients on views, clicks, orders, and GMV are large nor statistically significant, implying that PTG does not immediately impact how many products customers view or purchase, nor the price of those purchases. Note that this null effect may be due to at least two separate reasons. First, a customer navigating to the

[10] Limiting the analysis to customers who entered the experiment between mid March and mid July 2021 guarantees that we can track all those customers for at least 24 weeks.

[11] The total number of searches are the sum of PTG general searches, PTG specific searches, and non-PTG searches.

Table 2: Short-Run Impact on Number of Searches

| | Number of Searches | Number of PTG General Searches | Number of PTG Specific Searches |
|---|---|---|---|
| | (1) | (2) | (3) |
| Treat | 0.0486*** | 0.000479 | 0.0531*** |
| | (0.0152) | (0.00132) | (0.000628) |
| % Change | 1.04% | 0.04% | 495.22% |
| Observations | 505,485 | 505,485 | 505,485 |
| R-squared | 0.043 | 0.019 | 0.015 |

*Notes:* The dependent variables are in levels. % Change is calculated by dividing the treatment effect by the control group average. Standard errors are in parentheses. We include cohort fixed effects as per Equation 1.

e-commerce platform may indicate an underlying purchasing intent (say, buy a dress for a special event) that would not be affected by the availability of the PTG search tool. If this hypothesis were true, PTG may simply shift product views and purchases from one type of searches (non PTG or PTG general searches) to another (PTG specific searches). Second, making search more effective may have effects that take longer to materialize, as users learn to use the tool and to perform more effective searches on their own. We tackle the first hypothesis next, and the second hypothesis in the following sub-section.

Table 3: Short-Run Treatment Effects

| | Views | Clicks | Orders | GMV |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Treat | 0.666 | 0.0563 | 0.00537 | 0.35 |
| | (1.177) | (0.0465) | (0.0035) | (0.727) |
| %Change | 0.26% | 0.58% | 1.23% | 1.08% |
| Observations | 505,485 | 505,485 | 505,485 | 505,485 |
| R-squared | 0.013 | 0.025 | 0.008 | 0.002 |

*Notes:* The dependent variables are in levels. % Change is calculated by dividing the treatment effect by the control group average. Standard errors are in parentheses. We include cohort fixed effects as per Equation 1.

To evaluate whether PTG shifts consumers' viewing, clicking, and purchasing behavior toward PTG specific queries, we need to separate the effect on aggregate outcomes by query

type. We thus allocate product views, clicks, purchases, and expenditures to the three types of searches described in Section 3.1: PTG general, PTG specific, and non-PTG queries. We analyze the treatment effects on customer behavior for these three queries separately.

Results for PTG general and PTG specific queries are shown in Table 4.[12] The results confirm a significant and sizable decrease in the number of product views and clicks stemming from PTG general queries. Views decrease by 2.3%, and clicks decrease by 3.7%. Customers shift viewing and clicking to PTG specific queries. Columns 5 and 6 show that customers in the treatment group view 2.8 and click on 0.1 more products related to PTG specific queries compared to customers in the control group. Columns 7 and 8 further confirm that the shift in browsing behavior translates into 0.005 more products purchased and CNY0.255 more spent on products showing up in PTG specific queries. In percent terms, all these coefficients represent a more than 500% increase in the very small baseline browsing and purchasing behavior related to PTG specific queries. These estimates suggest that although the PTG search tool is not changing aggregate purchase intent, customers find the products they want through the help of narrower searches that are suggested by PTG.

Table 4: Decompose Short-Run Treatment Effects for PTG Queries

| | PTG General Queries | | | | PTG Specific Queries | | | |
|---|---|---|---|---|---|---|---|---|
| | Views | Clicks | Orders | GMV | Views | Clicks | Orders | GMV |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Treat | -1.402*** | -0.0846*** | -0.000989 | -0.412 | 2.756*** | 0.106*** | 0.00470*** | 0.255*** |
| | (0.277) | (0.0122) | (0.00117) | (0.269) | (0.0535) | (0.00227) | (0.0002) | (0.0225) |
| % Change | -2.34% | -3.74% | -0.80% | -4.69% | 515.3% | 516.59% | 568.36% | 560.19% |
| Observations | 505,485 | 505,485 | 505,485 | 505,485 | 505,485 | 505,485 | 505,485 | 505,485 |
| R-squared | 0.009 | 0.002 | 0.002 | 0.001 | 0.007 | 0.005 | 0.002 | 0.001 |

*Notes:* The dependent variables are in levels. % Change is calculated by dividing the treatment effect by the control group average. Standard errors are in parentheses. We include cohort fixed effects as per Equation 1. Appendix Table A.1 contains the estimates for non-PTG queries.

Overall, the results in this sub-section suggest that, while the PTG search tool impacts customer search behavior (as evidenced by the type of searches they conduct and the resulting purchases), in the immediate short-term it does not impact how many products customers buy or how much they spend on the platform.[13]

In order to understand whether the short-run results are due to PTG being ineffective at improving search or the customers needing time to find PTG valuable and learn from it,

[12] Results for non-PTG queries are presented in Appendix Table A.1.

[13] Appendix Table A.2 suggests that even if consumers buy different products in the treatment group relative to the control group, customer satisfaction does not change.

in the next sub-section we focus on the effects of the search tool in the long-run.

## 4.2 Long-Run Results

To investigate the long-run effects of introducing the PTG search tool, we restrict attention to 68% of users who entered the experiment early enough to give us about 6 months of experimental data for all of them (346,110 users). Specifically, we focus on customers who searched for PTG-related queries between mid March and July 16, 2021. This constraint ensures that we can observe all these customers for a minimum of 24 weeks by the end of 2021. Just like Table 2 in the previous sub-section, Appendix Table A.3 (column 3) confirms that PTG was effective at shifting customers in the treatment group to perform narrower searchers that were recommended by the tool itself: the number of PTG specific searches increases by 450%, off of an otherwise small baseline, whereas the total number of queries does not meaningfully change.

Given the impact of PTG on customer search behavior, we start by estimating the effects of the search tool on product views, clicks, purchases, and expenditures. We compute those outcomes at the consumer level by aggregating views, clicks, purchases, and expenditures over the course of the 24 weeks immediately following the customer's entry into the experiment.

Table 5 shows the estimates of Equation 1, where the outcomes are measured in the long-run. Starting from columns 1 and 2, the estimates imply that access to the new search tool does not significantly change the number of products viewed, nor the number of clicks on those products. Both the point estimates and the percentage changes are fairly small in magnitude. Columns 3 and 4 however, indicate that consumers with the PTG search tool purchase on average 0.34 more orders and spend on average CNY62 more compared to consumers without the search tool, an increase of 1.6% in orders and 3.2% in spending compared to the baseline. Together with the null results on product views and clicks (and the null effect on total searches from Appendix Table A.3), this purchase expansion seems primarily due to searches becoming more efficient in the treatment group, rather than customers dedicating more time to viewing and clicking on more products.

Given the null effects on purchasing behavior in the short-run (same-day) and the large positive effects in the long-run (the following 24 weeks), we want to explore how early these positive effects start emerging. Typically, A/B experiments are run for a few weeks, so this exercise can help us understand the extent to which short-run A/B experiments can capture the effects of product changes like ours.

To evaluate how early the positive effects on purchasing behavior materialize, we replicate columns 3 and 4 from Table 5 with outcomes aggregated over the first week, the first 2 weeks,

Table 5: Long-Run Treatment Effects

| | Views | Clicks | Orders | GMV |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Treat | -35.78 | -0.932 | 0.336** | 62.44** |
| | (54.09) | (2.215) | (0.140) | (27.60) |
| % Change | -0.29% | -0.19% | 1.57% | 3.24% |
| Observations | 346,110 | 346,110 | 346,110 | 346,110 |
| R-squared | 0.06 | 0.08 | 0.03 | 0.006 |

*Notes:* The dependent variables are in levels. % Change is calculated by dividing the treatment effect by the control group average. Standard errors are in parentheses. We include cohort fixed effects as per Equation 1.

the first 3 weeks, and the first 4 weeks since a customer enters the experiment.[14] Table 6 presents the results. Columns 1 and 2 report treatment effects on orders and expenditures in the first week following entry into the experiment. Columns 3 and 4 do the same for orders and expenditures in the first two weeks, and so on. All coefficients are statistically indistinguishable from 0, except for those in columns 7 and 8, which aggregate customer activity within the first month since entry into the experiment. In percentage terms, those effects (1.29% and 3.55%) are comparable to the longer-run results from Table 5, providing support for the hypothesis that it takes time for the positive benefits of improved search to materialize. The introduction of PTG progressively influences the purchasing decisions of the treatment group customers, ultimately leading to a significant increase in product orders and purchases. Despite the benefits, typical durations of A/B experiments would not be able to capture these benefits.[15]

In the rest of this section, we explore how the effects differ across consumers and products, we offer some evidence that the tool is effective at training users to perform better searches, and finally, we show that customers are more satisfied with the purchases they make, as evidenced by higher ratings and lower returns.

**Heterogeneous Effects.** Although PTG increases purchases on average, we are interested in exploring for which types of customers it is most effective in helping them find what

[14] Appendix Table A.4 presents analyses for time aggregations beyond the first four weeks. Those other time aggregations are all comparable to the long-run results presented in Table 5, obviously not in magnitudes, but rather in percentage terms.

[15] Note that the estimates in Table 6 and Appendix Table A.4 do not exactly resemble the analysis one could conduct with a short-run experiment. Indeed, the number of experimental participants, and hence the statistical power of the tests, is large only because the experiment was run over many months. In a sense, the analysis presented here offers an upper bound of what can be inferred from a short-run experiment.

Table 6: Treatment Effects Over Different Time Aggregations

| | Week 1 | | Weeks 1-2 | | Weeks 1-3 | | Weeks 1-4 | |
|---|---|---|---|---|---|---|---|---|
| | Orders | GMV | Orders | GMV | Orders | GMV | Orders | GMV |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Treat | 0.0157 | 2.758 | 0.0201 | 2.899 | 0.0310 | 7.366 | 0.0538* | 13.10** |
| | (0.0109) | (2.117) | (0.0174) | (3.480) | (0.0238) | (4.757) | (0.0302) | (5.964) |
| % Change | 1.12% | 2.34% | 0.86% | 1.42% | 0.95% | 2.57% | 1.29% | 3.55% |
| Observations | 346,110 | 346,110 | 346,110 | 346,110 | 346,110 | 346,110 | 346,110 | 346,110 |
| R-squared | 0.014 | 0.003 | 0.019 | 0.004 | 0.022 | 0.004 | 0.024 | 0.005 |

*Notes:* The dependent variables are in levels. In column 1, the number of orders placed by a customer is aggregated over the first week since the customer enters the experiment. In column 3, the number of orders is aggregated over the first two weeks, then in column 5 it is calculated over the first three weeks, and in columns 7 over the first four weeks. Columns 2, 4, 6, and 8 compute the same aggregations for expenditures. % Change is calculated by dividing the treatment effect by the control group average. Standard errors are in parentheses. We include cohort fixed effects as per Equation 1. Appendix Table A.4 presents similar analyses of longer time periods.

they want. To do this, we examine how the treatment effects on orders and spending differ among different customer categories. We pick five dimensions. Beyond gender, the other four dimensions proxy for how Internet savvy customers are: age, city of residence, year of registration on the platform, and frequency of platform use. For each of the four dimensions, we divide the customers into two separate groups and interact the treatment dummy with a variable denoting one of the two groups.

Results are presented in Table 7. Column 1 interacts the treatment indicator with a dummy for whether the customer is under 35 years old (58% of the experiment users are under 35 years old). The results indicate that the benefits of the search tool are concentrated among customers over 35 years old, providing support for the hypothesis that younger users already know how to effectively search for products online. The next three columns do not find support for heterogeneous effects. Column 2 interacts the treatment indicator with a dummy for whether the customer resides in a large city in China (32% of the experiment users live in a large city). Column 3 interacts the treatment indicator with a dummy for whether the customer is new to the platform, ie, they created their account in the last 5 years (40% of the users are considered new). Column 4 interacts the treatment indicator with a female dummy (55% of the customers self-identify as female). Finally, the last column shows a perhaps unexpected result, that more frequent users, despite their deeper knowledge of the platform, are those that truly benefit from PTG. Here, frequent users are defined as those in the top quartile of spending in the 8 weeks preceding the experiment.

Because PTG facilitates more specific and narrower searches, it is possible for it to affect the type of products consumers find. In particular, it is likely that PTG allows customers to

Table 7: Heterogeneous Treatment Effects on GMV (Customers)

| | GMV | GMV | GMV | GMV | GMV |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Treat | 117.5*** | 50.33 | 36.49 | 47.01 | 22.87 |
| | (41.70) | (32.55) | (34.67) | (40.60) | (30.95) |
| Treat*Young | -101.0* | | | | |
| | (54.44) | | | | |
| Treat*Big City | | 24.43 | | | |
| | | (57.37) | | | |
| Treat*New | | | 53.97 | | |
| | | | (54.60) | | |
| Treat*Female | | | | 19.82 | |
| | | | | (54.06) | |
| Treat*Heavy | | | | | 141.3** |
| | | | | | (61.90) |
| Observations | 346,110 | 346,110 | 346,110 | 346,110 | 346,110 |
| R-squared | 0.062 | 0.062 | 0.062 | 0.062 | 0.062 |

*Notes:* The dependent variables are GMV in levels. Column 1 reports different treatment results by customer age, where "Young=1" refers to consumers younger than 35 years old. Column 2 reports the results by city tier, where "Big City=1" denotes consumers residing in first and second-tier cities (in China, the cities are categorized into six tiers, and first and second-tier cities typically refer to large cities). Column 3 reports the results by the number of registered years on the platform, where "New=1" refers to consumers who created their account in the last five years. Column 4 reports the results by gender, where "Female=1" refers to consumers who self-identify as female. Column 5 reports the results by spending, where "Heavy=1" denotes consumers in the top quartile of expenditures during the 8 weeks prior to entering the experiment. Standard errors are in parentheses. We include cohort fixed effects as per Equation 1. The results of a similar analysis using Orders as the outcome variable are presented in Appendix Table A.5.

find less popular products. To test this hypothesis, we create two classifications of products into more versus less popular.

First, we compute product-level revenues for 2021,[16] which allows us to rank products from best to worst selling within their respective product categories. We then classify the products into five groups: top 10 selling products, the next 10-100 products, the next 100-1,000 products, the next 1,000-10,000 products, and finally, the products beyond the top 10,000 selling products. We then compare the expenditures of treatment and control customers in each of these five product groups.

Panel I of Table 8 shows the results. We find that products further down in the sales rank benefit the most because the tool makes it easier for consumers to find them. Columns 4 and 5 confirm that expenditures on products beyond the top 1,000 selling products increase by between 4.5% and 6% when PTG is available. For more popular products, the percentage increase is smaller (e.g., column 3) and sometimes even indistinguishable from zero (columns 1 and 2).

The second approach is to classify products by their seller's overall sales rank. We categorize sellers into five quantiles based on their cumulative annual revenues by the start of the roll out of PTG, and conduct a similar analysis. Panel II of Table 8 shows the results. The findings align closely with Panel I. Expenditures on top sellers show no significant change between treatment and control customers, perhaps because of the ease with which top sellers can already be found on the platform. In contrast, the two bottom quantiles of sellers experience a significant increase in revenues from treated customers, by between 3.7% and 5%. Together, the two panels of Table 8 indicate that tail and niche products gain more from the introduction of PTG.

**Mechanisms: Search Tool Suggestions and Consumer Learning.** There are a number of possible mechanisms explaining the aggregate improvement in search effectiveness that we find, i.e., the improvement in purchases and expenditures given a comparable number of product views and clicks presented in Table 5. The first possibility is that the search tool allows customers to refine their searches whenever the tool is available. This is the most direct benefit of the search tool, which would imply that the improvements are concentrated on searches where the tool is available, i.e. PTG general and specific queries.

To test this, columns 1 and 2 in Table 9 estimate the effect of the search tool on orders and expenditures generated through PTG general and specific queries. The coefficient estimates confirm large and significant effects of the search tool for PTG queries: orders increase by

[16]The purchases related to the users in our experiment are a small share of the revenues for these products in 2021, so the likelihood that our experimental treatment affects the product sales rank is very low.

Table 8: Heterogeneous Treatment Effects on GMV (Products and Sellers)

| | Panel I: GMV Across Products Grouped by Products' GMV Rank | | | | |
|---|---|---|---|---|---|
| | Top 10 | 10-100 | 100-1000 | 1000-10000 | Beyond 10000 |
| | (1) | (2) | (3) | (4) | (5) |
| Treat | 1.639 | 5.668 | 15.25* | 20.60** | 19.57*** |
| | (7.208) | (6.382) | (9.097) | (8.027) | (5.954) |
| % Change | 0.77% | 1.55% | 2.89% | 4.53% | 5.96% |
| Observations | 346,110 | 346,110 | 346,110 | 346,110 | 346,110 |
| R-squared | 0.001 | 0.005 | 0.004 | 0.004 | 0.004 |
| | Panel II: GMV Across Products Grouped by Sellers' Revenue Quantile | | | | |
| | Top 20% | Med-high 20% | Medium 20% | Med-low 20% | Tail 20% |
| | (1) | (2) | (3) | (4) | (5) |
| Treat | 3.120 | 3.473 | 15.13 | 26.05** | 14.96** |
| | (2.857) | (6.906) | (10.27) | (10.23) | (7.078) |
| % Change | 2.03% | 1.03% | 3.17% | 5.00% | 3.74% |
| Observations | 346,110 | 346,110 | 346,110 | 346,110 | 346,110 |
| R-squared | 0.005 | 0.004 | 0.003 | 0.003 | 0.004 |

*Notes:* The dependent variables, always GMV, are in levels. In Panel I, products are classified into 5 groups, depending on their GMV rank within their respective product categories. In Panel II, products are classified into another 5 groups, depending on their sellers' revenue rank. % Change is calculated by dividing the treatment effect by the control group average. Standard errors are in parentheses. We include cohort fixed effects according as per Equation 1. The results of a similar analysis using Orders as the outcome variable are presented in Appendix Table A.6.

2.2% and expenditures increase by 5.3% in the 24 weeks since the customer's entry in the experiment.

Table 9: Treatment Effects for PTG and Non-PTG Queries

| | PTG Queries | | Non-PTG Queries | |
|---|---|---|---|---|
| | Orders | GMV | Orders | GMV |
| | (1) | (2) | (3) | (4) |
| Treat | 0.0453*** | 7.779*** | 0.291** | 54.67** |
| | (0.0128) | (2.452) | (0.131) | (26.00) |
| % Change | 2.22% | 5.34% | 1.50% | 3.07% |
| Observations | 346,110 | 346,110 | 346,110 | 346,110 |
| R-squared | 0.01 | 0.001 | 0.031 | 0.006 |

*Notes:* The dependent variables are in levels. % Change is calculated by dividing the treatment effect by the control group average. Standard errors are in parentheses. We include cohort fixed effects as per Equation 1.

The second, more indirect, channel through which the search tool can be helpful is by increasing overall customer satisfaction with the platform, which in turn can increase customer loyalty. This hypothesis would imply that customers in the treatment group use the platform more often than customers in the control group. To test this, we consider two metrics: the number of days performing searches and the number of search query categories.[17] We regress the two metrics on the treatment indicator to test whether treated customers use the platform more than control customers. As shown in Appendix Table A.7, the PTG search tool does not increase the overall usage of the platform.

Finally, the third possibility that we explore is that customers learn to perform more effective searches from PTG, which they can then apply even to product searches that are not augmented by PTG. If this were the case, we would expect an increase in orders and spending originating from non-PTG searches, as well as a shift of non-PTG queries towards finer and more specific searches.

We test whether the tool leads to an increase in orders and spending from non-PTG queries. Columns 3 and 4 in Table 9 confirm sizable effects: orders increase by 1.5% and spending increases by 3.1%. Although in levels, the increases estimated in columns 3 and 4 are larger than those estimated in columns 1 and 2 (because a bigger share of sales come from non-PTG queries), in percentage terms the opposite is true. This result confirms that

[17]As each query is related to various purchased products, we can calculate each query's number of orders within every product category. We define a query's category as the one with the highest number of orders attributed to it.

in percentage terms, the direct effect of the search tool on PTG queries (columns 1 and 2) is larger than the indirect effect on non-PTG queries (columns 3 and 4).

Does the increase in orders and spending on non-PTG queries come from customers learning to perform more effective searches on their own? We start to explore this possibility by focusing on the text length of customer searches. We would expect that customers in the treatment group may learn to conduct more specific searches, perhaps using longer descriptions of the items they want. We measure query length as the number of Chinese characters that the customer types in the search box. We then compute the average query length across all searches performed in the 24 weeks since entry in the experiment, and the average query length across all non-PTG searches.

Table 10: Tests for Customer Learning

| Outcome: | Avg Query Length | Avg Query Length | Number of Searches | Number of Searches |
|---|---|---|---|---|
| Query Type: | All Queries | Non-PTG | Matched Non-PTG | Unmatched Non-PTG |
| | (1) | (2) | (3) | (4) |
| Treat | 0.0191*** | 0.0126** | 0.313** | -0.790 |
| | (0.00479) | (0.00535) | (0.133) | (0.682) |
| % Change | 0.31% | 0.20% | 2.41% | -0.40% |
| Observations | 346,110 | 346,110 | 346,110 | 346,110 |
| R-squared | 0.021 | 0.045 | 0.012 | 0.135 |

*Notes:* The dependent variables are in levels. % Change is calculated by dividing the treatment effect by the control group average. Standard errors are in parentheses. We include cohort fixed effects according as per Equation 1.

Columns 1 and 2 of Table 10 present the results. Across all searches, the average query length increases by 0.02 characters, or 0.3%. For non-PTG searches, the result is smaller in magnitude, as expected given that it is an indirect effect, but statistically different from zero at conventional levels.

As they search for products not included in PTG, customers might also start using words they learned from PTG queries and integrate them into other searches. For instance, a consumer could discover the term "Halterneck" from a PTG specific query when searching for dresses, and then use the same descriptor in other searches, e.g., "Halterneck Top." To investigate this possibility, we make a list of all the words that the search tool uses in PTG specific queries (e.g., "Halterneck"). We call this list the PTG specific vocabulary. We then categorize non-PTG queries into two groups: queries whose words match at least one word

included in the PTG specific vocabulary (*matched non-PTG queries*); and queries whose words do not match any of the words included in the PTG specific vocabulary (*unmatched non-PTG queries*). In the example above, the search for "Halterneck Top" would be classified as a matched non-PTG query. We want to test whether the number of matched non-PTG queries increases in the treatment group relative to the control group, potentially at the expense of unmatched queries.

Columns 3 of Table 10 shows that the number of matched non-PTG queries significantly increases by 0.3 queries, or 2.4% relative to the baseline level. Although the coefficient estimate is negative in column 4, the increase in matched queries does not seem to come at the expense of unmatched queries.

Table 11: Effects on Customer Satisfaction: Positive Reviews and Return Rates

| | Positive Rating | Return |
|---|---|---|
| | (1) | (2) |
| Treat | 0.00636*** | -0.00274*** |
| | (0.000201) | (0.0000731) |
| % Change | 3.56% | -3.22% |
| Observations | 7,479,300 | 7,479,300 |
| R-squared | 0.003 | 0.012 |

*Notes:* Linear probability estimates of Equation 2. An observation is an order placed in the 24 weeks following a customer's entry in the experiment. The dependent variables are indicator for a positive rating (column 1) and request for return and refund (column 2). Standard errors are in parentheses. Similar regressions restricting attention to orders related to PTG queries are displayed in Appendix Table A.8.

**Customer Satisfaction.** So far we have showed that consumers perform more effective searches through the help of PTG. The benefits arise both from the direct use of the tool, and the indirect learning that the tool provides. Before concluding, we want to ensure that the additional purchases induced by PTG are as good as or better than the purchases customers would make in the absence of PTG.

We consider all orders placed in the 24 weeks since the customer's entry into the experiment. This amounts to 7,479,300 orders placed by the 346,110 customers for whom we have at least 24 weeks of data, for an average of 22 orders per customer. We run two linear probability models of the following type:

$$y_{ij} = \beta * Treat_i + \alpha_{c(i)} + \epsilon_{ij}, \tag{2}$$

where $i$ denotes the costumer as in Equation 1, and $j$ denotes a purchased product order during the relevant time period. For $y_{ij}$, we use two proxies for customer satisfaction: an indicator for whether the customer submits a positive review for the purchased product (i.e., 4- or 5-star review out of a 1-5 scale); and an indicator for whether the customer returns the product and requests a refund. We also control for customers' day of entry into the experiment with cohort $c(i)$ fixed effects.

Results are presented in Table 11. Purchases from customers in the treatment group are substantially more likely to be rated positively (0.64 percentage points more likely) and less likely to be returned (0.27 percentage points less likely). The effects are substantial relative to the baseline, with a 3.56% increase in the positive rating probability and a 3.22% decrease in the return probability. The results thus confirm that the purchases made with the support of PTG are perceived as higher quality.

Together, our results showcase the importance for platforms to keep on improving search design. Search refinement tools such as the one we study here can be effective not just directly, but also indirectly, by teaching customers to more effectively discover and describe their preferences.

# 5 Conclusion

One of the most important roles of digital platforms is to facilitate matches between many buyers and sellers of products and services. By developing increasingly sophisticated ranking algorithms, platforms have invested substantial efforts in making search results as relevant as possible given what consumers say they want. However, less emphasis has been put on helping consumers identify and effectively describe what they want.

Our research shows that improving search tools with textual and visual suggestions can be an effective way to help consumers express and develop their preferences. Leveraging a long-run experiment linked to the launch of a text- and picture-based search refinement tool on a major e-commerce platform, we find that having access to textual and visual search recommendations increases purchases by 1.57% and spending by 3.24%. The increase does not seem to be driven by consumers viewing or clicking on more products, nor by consumers conducting more searches, suggesting an increase in search effectiveness. The increase is not only driven by searches that are directly affected by the search tool, but rather it spills over to other searches, implying that consumers learn to perform better searches on their own.

The value of the tool is concentrated among a subset of buyers and a subset of sellers. On the buyer side, we find two distinct results. On one hand, the tool helps older consumers find what they need. Younger consumers do not seem to be greatly affected by the search

refinement tool, perhaps given their intrinsic ability to search online. On the other hand, we also find that heavier consumers of the platform (rather than lighter consumers) benefit the most from the tool. The latter result suggests that experience with the platform is not enough to reduce search costs related to demand expression and demand formation.

Although the experiment directly affected demand, we find important indirect effects for sellers as well. In particular, the ability of the search tool to narrow down searches to subsets of more specific categories benefits products and sellers outside of the most popular, reducing the concentration of sales among the top sellers.

Importantly, we also find that if we restrict our analysis to the short-run, we are unable to detect the significant benefits of the new search tool. In fact, we find that the immediate effect on the first day a customer enters the experiment is a precisely estimated zero. This result likely reflects the fact that people entering the experiment are visiting the platform with a specific intent to buy (or not to buy) something, which the search tool does not immediately impact. It also implies that it takes time for search tools like the one we study to display their beneficial effects on consumers. In this specific case, our analysis reveals that it takes about one month since entry in the experiment for consumers to experience the improvements in search effectiveness driven by the search refinement tool.

Our results have two important implications for the design of search mechanisms and for the design of experiments. Related to search mechanisms, our results highlight the importance of addressing demand expression and demand formation challenges. The design of search mechanisms is evolving away from purely relying on consumer prompts to identify what consumers want, towards increasing the role of machine learning to leverage large data on consumer search and purchasing behavior. Search recommendation tools can serve at least two main purposes: refining choices to a subset of the options available; or expanding choices to a superset of the options available, potentially including both substitute and complementary options. Our refinement tool clearly falls in the first group, which is particularly valuable in a context where too many options can make choosing harder. However, since many recommendation tools are designed to provide a combination of refinement and expansion opportunities, separately identifying when and how they can make search easier is an important question for future investigation.

Search mechanisms have historically been relying primarily on textual prompts. Only in recent years, some digital platforms have started offering visual recommendations (such as Amazon) or visual-based searches (such as Google Lens). Although our research cannot separate the role of textual versus visual suggestions, an important avenue for future research would be to identify the separate and complementary benefits of the two.

Related to experiment design, our results highlight the risk of drawing conclusions from

short-run experiments. Despite recent efforts to identify long-run results from short-term metrics (Athey et al., 2019), there are many contexts, such as ours, where the best approach is simply to run a long-run experiment. In our case, if we had only had access to a few weeks of data, we would have concluded that the search tool did not make consumer search more effective, both because the estimates are too noisy to detect a significant effect and because the short-term impact is quantitatively close to zero. It is thus important to expand research to understand when and how practitioners and researchers can rely on short-run experiments, and when instead longer run approaches may be required.

The paper has a number of limitations. First, our analysis is unable to look at how sellers would respond to the roll out of search refinement tools like the one we study. There are two main reasons for that. Although increasing over the course of the experimental period, the searches qualifying for the refinement tool accounted for only about 15% of GMV linked to searches (Figure 2) by the end of 2021. Additionally, because of the experimental roll-out, not all users had the opportunity to benefit from the search tool recommendations. We leave the important question of how sellers would adjust their product offering in response to changes in search design to future research.

Second, we have robust analyses indicating that the tool had net benefits on consumers: they purchased more given the same search effort, and those purchases had higher ratings and lower returns. However, many recommendation systems run the risk of recommending impulse purchases that consumers may ex-post regret. It is possible that ratings and return rates may not capture this type of longer-term regret. Similarly, because choices may not reflect true preferences, search recommendation tools risk diverting consumers away from their true preferences. In the financial setting, for example, this has been identified as a potential risk of autocomplete tools for stock tickers Rubin and Rubin (2021). The combination of search aid tools, ranking algorithms, and product recommendations, Mik (2016) argues, risks eroding consumer autonomy in online transactions. In our setting, this is unlikely to happen given the immediate positive effects on two proxies for customer satisfaction – higher ratings and lower returns. However, it is important for research to test the potential costs of search aid and recommendations in nudging consumers to spend beyond their means, or to buy products that they would not otherwise want.

# Funding and Competing Interests

The first author acknowledges financial support from NSFC Grants 72192803.
The fourth author acknowledges financial support from NSFC Grants 72203202 and 72192803.
The fifth author is an employee of the collaborating platform.
The other authors have no funding or competing interests to disclose.

# References

**Adomavicius, Gediminas, Jesse Bockstedt, Shawn P Curley, Jingjing Zhang, and Sam Ransbotham**, "The hidden side effects of recommendation systems," *MIT Sloan Management Review*, 2019, *60* (2), 1.

**_ , Jesse C Bockstedt, Shawn P Curley, and Jingjing Zhang**, "Do recommender systems manipulate consumer preferences? A study of anchoring effects," *Information Systems Research*, 2013, *24* (4), 956–975.

**_ , _ , _ , and _** , "Effects of online recommendations on consumers' willingness to pay," *Information Systems Research*, 2018, *29* (1), 84–102.

**Athey, Susan, Raj Chetty, Guido W Imbens, and Hyunseung Kang**, "The surrogate index: Combining short-term proxies to estimate long-term treatment effects more rapidly and precisely," *NBER Working Paper No. 26463*, 2019.

**Bar-Isaac, Heski, Guillermo Caruana, and Vicente Cuñat**, "Search, design, and market structure," *American Economic Review*, 2012, *102* (2), 1140–1160.

**Blanco, Carlos Flavián, Raquel Gurrea Sarasa, and Carlos Orús Sanclemente**, "Effects of visual and textual information in online product presentations: looking for the best combination in website design," *European Journal of information systems*, 2010, *19* (6), 668–686.

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

**Brynjolfsson, Erik, Yu Hu, and Duncan Simester**, "Goodbye pareto principle, hello long tail: The effect of search costs on the concentration of product sales," *Management science*, 2011, *57* (8), 1373–1386.

**Chen, Jiafeng and Jonathan Roth**, "Logs with zeros? Some problems and solutions," 2023.

**Chen, Yuxin and Song Yao**, "Sequential search with refinement: Model and application with click-stream data," *Management Science*, 2019, *63* (12), 4345–4365.

**_ , Zhe Yuan, Tianshu Sun, and AJ Chen**, "Understanding the Impacts of Depersonalization in Search Algorithm on Consumer Behavior: A Field Experiment with a Large Online Retail Platform," *Available at SSRN 4412157*, 2023.

**Chiou, Lesley and Catherine Tucker**, "Search engines and data retention: Implications for privacy and antitrust," Technical Report, National Bureau of Economic Research 2017.

**Choi, Michael, Anovia Yifan Dai, and Kyungmin Kim**, "Consumer search and price competition," *Econometrica*, 2018, *86* (4), 1257–1281.

**Dinerstein, Michael, Liran Einav, Jonathan Levin, and Neel Sundaresan**, "Consumer price search and platform design in internet commerce," *American Economic Review*, 2018, *108* (7), 1820–1859.

**Elberse, Anita and Felix Oberholzer-Gee**, "Superstars and underdogs: An examination of the long tail phenomenon in video sales," *Division of Research, Harvard Business School*, 2006, *7*.

**Ellison, Glenn and Sara Fisher Ellison**, "Search, obfuscation, and price elasticities on the internet," *Econometrica*, 2009, *77* (2), 427–452.

**Filippas, Apostolos, John J Horton, and Diego Urraca**, "Advertising as Coordination: Evidence from a Field Experiment," *Working Paper*, 2023.

**Fleder, Daniel and Kartik Hosanagar**, "Blockbuster culture's next rise or fall: The impact of recommender systems on sales diversity," *Management Science*, 2009, *55* (5), 697–712.

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

**Fradkin, Andrey**, "Search, matching, and the role of digital marketplace design in enabling trade: Evidence from airbnb," *Matching, and the Role of Digital Marketplace Design in Enabling Trade: Evidence from Airbnb (March 21, 2017)*, 2017.

**Gardner, Martin**, "Mathematical games," *Scientific american*, 1970, *222* (6), 132–140.

**Goli, Ali, David H. Reiley, and Hongkai Zhang**, "Personalized Versioning: Product Strategies Constructed from Experiments on Pandora," *Working Paper*, 2021.

**Gupta, Somit, Ronny Kohavi, Diane Tang, Ya Xu, Reid Andersen, Eytan Bakshy, Niall Cardin, Sumita Chandran, Nanyu Chen, Dominic Coey et al.**, "Top challenges from the first practical online controlled experiments summit," *ACM SIGKDD Explorations Newsletter*, 2019, *21* (1), 20–35.

**Häubl, Gerald and Kyle B Murray**, "Preference construction and persistence in digital marketplaces: The role of electronic recommendation agents," *Journal of consumer psychology*, 2003, *13* (1-2), 75–91.

_ **and Valerie Trifts**, "Consumer decision making in online shopping environments: The effects of interactive decision aids," *Marketing science*, 2000, *19* (1), 4–21.

**Huang, Jason, David Reiley, and Nick Riabov**, "Measuring consumer sensitivity to audio advertising: A field experiment on pandora internet radio," *Available at SSRN 3166676*, 2018.

**Jiang, Baojun and Tianxin Zou**, "Consumer search and filtering on online retail platforms," *Journal of Marketing Research*, 2020, *57* (5), 900–916.

**Kamenica, Emir, Sendhil Mullainathan, and Richard Thaler**, "Helping consumers know themselves," *American Economic Review*, 2011, *101* (3), 417–422.

**Kohavi, Ron, Alex Deng, Brian Frasca, Roger Longbotham, Toby Walker, and Ya Xu**, "Trustworthy online controlled experiments: Five puzzling outcomes explained," in "Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining" 2012, pp. 786–794.

**Lazonder, Ard W**, "Do two heads search better than one? Effects of student collaboration on web search behaviour and search outcomes," *British journal of educational technology*, 2005, *36* (3), 465–475.

**Lee, Kwok Hao and Leon Musolff**, "Entry into two-sided markets shaped by platform-guided search," *Working Paper*, 2021.

**Lei, Xiaoxia, Yixing Chen, and Ananya Sen**, "The Value of External Data for Digital Platforms: Evidence from a Field Experiment on Search Suggestions," *Available at SSRN*, 2023.

**Liu, Jia and Olivier Toubia**, "Search query formation by strategic consumers," *Quantitative Marketing and Economics*, 2020, *18*, 155–194.

**los Santos, Babur De, Ali Hortaçsu, and Matthijs R Wildenbeest**, "Testing models of consumer search using data on web browsing and purchasing behavior," *American economic review*, 2012, *102* (6), 2955–2980.

**Markey, Karen**, *Online searching: A guide to finding quality information efficiently and effectively*, Rowman & Littlefield, 2019.

**Mik, Eliza**, "The erosion of autonomy in online consumer transactions," *Law, Innovation and Technology*, 2016, *8* (1), 1–38.

**Rothschild, Michael**, "Searching for the Lowest Price When the Distribution of Prices Is Unknown," *Journal of Political Economy*, 1974, *82* (4), 689–711.

**Rubin, Eran and Amir Rubin**, "On the economic effects of the text completion interface: empirical analysis of financial markets," *Electronic Markets*, 2021, *31* (3), 717–735.

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

**Sun, Tianshu, Zhe Yuan, Chunxiao Li, Kaifu Zhang, and Jun Xu**, "The Value of Personal Data in Internet Commerce: A High-Stakes Field Experiment on Data Regulation Policy," *Management Science*, 2023.

**Tong, S., E. Kwon, S. Zhang, and G. Burtch**, "Recommending What to Search: How a Query Recommender System Affects Mobile Shopping," *Working Paper*, 2023.

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

**Wan, Xiang (Shawn), Anuj Kumar, and Xitong Li**, "How Do Product Recommendations Help Consumers Search? Evidence from a Field Experiment," *Management Science*, 2023.

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

**Wu, Ruijuan, Heng-Hui Wu, and Cheng Lu Wang**, "Why is a picture 'worth a thousand words'? Pictures as information in perceived helpfulness of online reviews," *International Journal of Consumer Studies*, 2021, *45* (3), 364–378.

**Yang, Joonhyuk, Navdeep S Sahni, Harikesh S Nair, and Xi Xiong**, "Advertising as information for ranking e-commerce search listings," *Marketing Science*, 2023.

**Yuan, Zhe, AJ Chen, Yitong Wang, and Tianshu Sun**, "How Recommendation Affects Customer Search: A Field Experiment," *Available at SSRN*, 2023.

# A Additional Figures and Tables

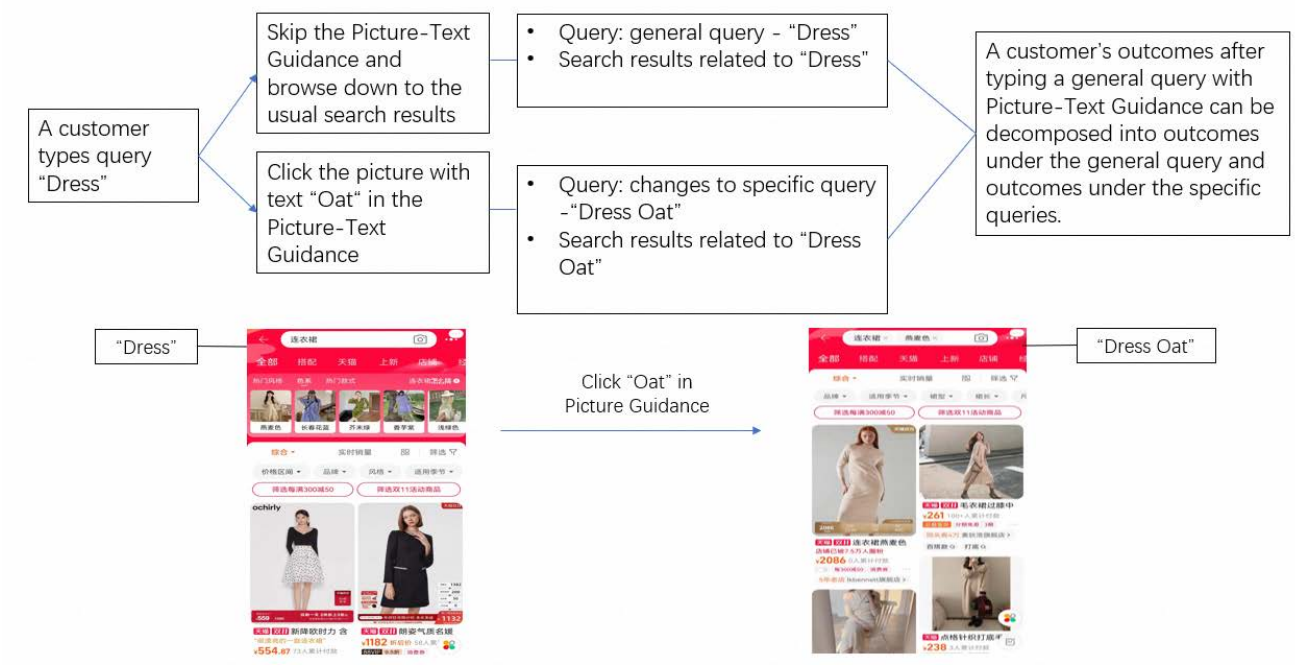



Figure A.1: Customer Search Process with Picture-Text Guidance

*Notes:* Treatment group customers have two options after searching for a PTG general query word. For example, if a customer types in "Dress" (PTG general query), they can either skip Picture-Text Guidance and search for products related to "Dress," or click on the picture with text "Oat" and search for products related to "Dress Oat" (PTG specific query) instead of just "Dress".

Table A.1: Short-Run Treatment Effects for Non-PTG Queries

| | Views | Clicks | Orders | GMV |
|---|---|---|---|---|
| | (1) | (2) | (3) | (4) |
| Treat | -0.687 | 0.0352 | 0.00166 | 0.507 |
| | (1.081) | (0.0427) | (0.0031) | (0.638) |
| % Change | -0.35% | 0.47% | 0.53% | 2.14% |
| Observations | 505,485 | 505,485 | 505,485 | 505,485 |
| R-squared | 0.018 | 0.029 | 0.01 | 0.002 |

*Notes:* The table presents results similar to Table 4. Instead of focusing on PTG queries, it restrictions attention to views, clicks, orders, and GMV associated to non-PTG queries.

Table A.2: Short-Run Effects on Customer Satisfaction: Positive Reviews and Return Rates

| | Positive Rating | Return |
|---|---|---|
| | (1) | (2) |
| Treat | 3.82e-07 | -0.00115 |
| | (0.00137) | (0.00094) |
| % Change | 0.0003% | -1.37% |
| Observations | 222,517 | 222,517 |
| R-squared | 0.006 | 0.014 |

*Notes:* An observation is a completed purchase on the day a customer enters the experiment. The dependent variable in column 1 is an indicator for whether the customer leaves a 4-star or 5-star rating for the purchase. The dependent variable in column 2 is an indicator for whether the customer requests a return of the item. Standard errors are in parentheses. We include cohort fixed effects.

Table A.3: Long-Run Impact on Number of Searches

| | Number of Searches | Number of PTG General Searches | Number of PTG Specific Searches |
|---|---|---|---|
| | (1) | (2) | (3) |
| Treat | 0.172 | -0.0212 | 0.670*** |
| | (0.814) | (0.0583) | (0.00523) |
| % Change | 0.08% | -0.14% | 450.55% |
| Observations | 346,110 | 346,110 | 346,110 |
| R-squared | 0.116 | 0.034 | 0.048 |

*Notes:* The dependent variables are in levels. % Change is calculated by dividing the treatment effect by the control group average. Standard errors are in parentheses. We include cohort fixed effects according to equation 1.

Table A.4: Treatment Effects Over Different Time Aggregations (for 8, 12, 16, 20 weeks)

| | Weeks 1-8 | | Weeks 1-12 | | Weeks 1-16 | | Weeks 1-20 | |
|---|---|---|---|---|---|---|---|---|
| | Orders | GMV | Orders | GMV | Orders | GMV | Orders | GMV |
| | (1) | (2) | (3) | (4) | (5) | (6) | (7) | (8) |
| Treat | 0.107* | 20.59* | 0.161** | 31.58** | 0.202** | 41.89** | 0.269** | 53.32** |
| | (0.0563) | (10.93) | (0.0791) | (15.55) | (0.101) | (19.79) | (0.122) | (23.92) |
| % Change | 1.38% | 2.94% | 1.44% | 3.11% | 1.39% | 3.18% | 1.50% | 3.30% |
| Observations | 346,110 | 346,110 | 346,110 | 346,110 | 346,110 | 346,110 | 346,110 | 346,110 |
| R-squared | 0.028 | 0.006 | 0.030 | 0.006 | 0.029 | 0.006 | 0.029 | 0.006 |

*Notes:* The dependent variables are in levels. In column 1, the number of orders placed by a customer is aggregated over the first eight weeks since the customer enters the experiment. In column 3, the number of orders is aggregated over the first twelve weeks, then in column 5 it is calculated over the first sixteen weeks, and in columns 7 over the first twenty weeks. Columns 2, 4, 6, and 8 compute the same aggregations for expenditures. % Change is calculated by dividing the treatment effect by the control group average. Standard errors are in parentheses. We include cohort fixed effects as per Equation 1.

Table A.5: Heterogeneous Treatment Effects on Orders (Customers)

| | Orders | Orders | Orders | Orders | Orders |
|---|---|---|---|---|---|
| | (1) | (2) | (3) | (4) | (5) |
| Treat | 0.281 | 0.234 | 0.216 | 0.392** | 0.159 |
| | (0.205) | (0.160) | (0.170) | (0.199) | (0.152) |
| Treat*Young | 0.061 | | | | |
| | (0.267) | | | | |
| Treat*Big City | | 0.258 | | | |
| | | (0.281) | | | |
| Treat*New | | | 0.251 | | |
| | | | (0.268) | | |
| Treat*Female | | | | -0.134 | |
| | | | | (0.265) | |
| Treat*Heavy | | | | | 0.632** |
| | | | | | (0.304) |
| Observations | 346,110 | 346,110 | 346,110 | 346,110 | 346,110 |
| R-squared | 0.146 | 0.146 | 0.146 | 0.146 | 0.146 |

*Notes:* The dependent variables are Orders in levels. Column 1 reports different treatment results by customer age, where "Young=1" refers to consumers younger than 35 years old. Column 2 reports the results by city tier, where "Big City=1" denotes consumers residing in first and second-tier cities (in China, the cities are categorized into six tiers, and first and second-tier cities typically refer to large cities). Column 3 reports the results by the number of registered years on the platform, where "New=1" refers to consumers who created their account in the last five years. Column 4 reports the results by gender, where "Female=1" refers to consumers who self-identify as female. Column 5 reports the results by spending, where "Heavy=1" denotes consumers in the top quartile of expenditures during the 8 weeks prior to entering the experiment. Standard errors are in parentheses. We include cohort fixed effects as per Equation 1.

Table A.6: Heterogeneous Treatment Effects on Orders (Products and Sellers)

| | Panel I: Orders Across Products Grouped by Products' GMV Rank | | | | |
|---|---|---|---|---|---|
| | Top 10 | 10-100 | 100-1000 | 1000-10000 | Beyond 10000 |
| | (1) | (2) | (3) | (4) | (5) |
| Treat | 0.0182 | 0.0438** | 0.0653 | 0.0863** | 0.119** |
| | (0.0125) | (0.0216) | (0.0403) | (0.0399) | (0.0527) |
| % Change | 0.8% | 1.12% | 1.23% | 1.78% | 2.59% |
| Observations | 346,110 | 346,110 | 346,110 | 346,110 | 346,110 |
| R-squared | 0.047 | 0.046 | 0.024 | 0.018 | 0.01 |
| | Panel II: Orders Across Products Grouped by Sellers' Revenue Quantile | | | | |
| | Top 20% | Med-high 20% | Medium 20% | Med-low 20% | Low 20% |
| Treat | 0.0113 | 0.0176 | 0.0740** | 0.113*** | 0.118*** |
| | (0.0181) | (0.0441) | (0.0361) | (0.0369) | (0.0452) |
| % Change | 0.48% | 0.44% | 1.62% | 2.31% | 2.31% |
| Observations | 346,110 | 346,110 | 346,110 | 346,110 | 346,110 |
| R-squared | 0.029 | 0.013 | 0.023 | 0.022 | 0.014 |

*Notes:* The table is identical to Table 8 except that the outcome variable is number of Orders rather than GMV. The dependent variable is in levels in all columns. % Change is calculated by dividing the treatment effect by the control group average. Standard errors are in parentheses. We include cohort fixed effects according as per Equation 1.

Table A.7: Effects on Customer Frequency of Use of the Platform

| | Number of Search Days | Number of Query Categories |
|---|---|---|
| | (1) | (2) |
| Treat | -0.099 | -0.039 |
| | (0.120) | (0.290) |
| % Change | -0.16% | -0.04% |
| Observations | 346,110 | 346,110 |
| R-squared | 0.14 | 0.14 |

*Notes:* The dependent variables are in levels. This table considers two metrics: the number of days performing searches and the number of search query categories. As each query is related to various purchased products, we can calculate each query's number of orders within every product category. We define a query's category as the one with the highest number of orders attributed to it. % Change is calculated by dividing the treatment effect by the control group average. Standard errors are in parentheses. We include cohort fixed effects as per Equation 1.

Table A.8: Positive Reviews and Return Rates for PTG queries

| | Positive Rating | Return |
|---|---|---|
| | (1) | (2) |
| Treat | 0.00515*** | -0.00363*** |
| | (0.000686) | (0.000384) |
| % Change | 3.28% | -4.58% |
| Observations | 713,883 | 713,883 |
| R-squared | 0.002 | 0.004 |

*Notes:* The table is identical to Table 11, except that the observations are restricted to orders directly related to PTG queries.