Platform Information Provision and Consumer Search: A Field Experiment*

Yanyou Chen
University of Toronto

Lu Fang
Zhejiang University

Chiara Farronato
Harvard University, CEPR, and NBER

Zhe Yuan
Zhejiang University

This Version: *October 23, 2023*

Abstract

Despite substantial efforts to help consumers search online in more intuitive ways, text search remains the predominant tool for product discovery online. In this paper, we explore the effects of visual cues on consumer search and purchasing behavior. We collaborate with one of 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 specific queries (e.g., "bluetooth headphones" or "noise-canceling headphones") with the help of images and texts to narrow down the search. 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, in the long-run, the search tool changed customers' search and purchasing behavior. In particular, 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. We find that the effect is not only driven by the direct value of suggested searches. Indeed, we find evidence that consumers learn to perform more effective searches across many product categories.

Keywords: Two-Sided Markets, Matching, Digital Platforms, Consumer Search, Learning, Search Recommendations

*Contact: Chen: yanyou.chen@utoronto.ca; Fang: corresponding author, fl_fanglu@163.com; Farronato: cfarronato@hbs.edu; Yuan: yyyuanzhe@gmail.com. Any opinions, findings, conclusions and recommendations expressed are those of the authors and do not necessarily represent the views of the focal platform. We are grateful to Avi Goldfarb, Joshua Gans, Heski Bar-Isaac, and seminar participants at the University of Toronto and Zhejiang University for helpful discussions and comments. All errors are our own.

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, we have 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 spend 3.2% more and complete 1.6% more orders compared to the control group. We find evidence that this increase in consumer activity 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, whether or not they are aided by the new tool. 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 results have implications for the design of search mechanisms online. On one hand, consumers have private information over what they want to search online. On the other, they

[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/

may benefit from recommendations to refine their searches and from inspirations for products that they may not ex-ante know they want. Our results suggest that the current design of search mechanisms may still rely too much on consumers' prompts given the extensive knowledge that platforms have about consumer preferences, in the aggregate as well as at the individual level.

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

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), as well as quantifications of search frictions (Lee and Musolff, 2021; Ellison and Ellison, 2009). More recently, Choi et al. (2018) focus on the unexpected consequences of lowering search costs.

In the context of eBay, Dinerstein et al. (2018) is one of the earliest works that looks at how platform search design plays a critical role in reducing search frictions and changing competition, ultimately determining market outcomes and welfare. 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.

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).

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 recommendations (or 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. Chen et al. (2023) conduct a field experiment on a search engine, investigating the effects of de-personalized search algorithms on consumer search, finding that de-personalization leads to longer browsing sessions, lower click rates, and lower purchase rates.

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, which to date is limited, by highlighting the role of visual and textual suggestions in guiding consumer search.

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.

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. 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 or browsing through online platforms. 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 supports the hypothesis 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.

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 in 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

[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.

them later.

The focus of our study is a new search tool that suggests a combination of picture and text options for the consumer to refine their searches. We call this tool *PT guidance* (or PTG, short for Picture-Text guidance) 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.

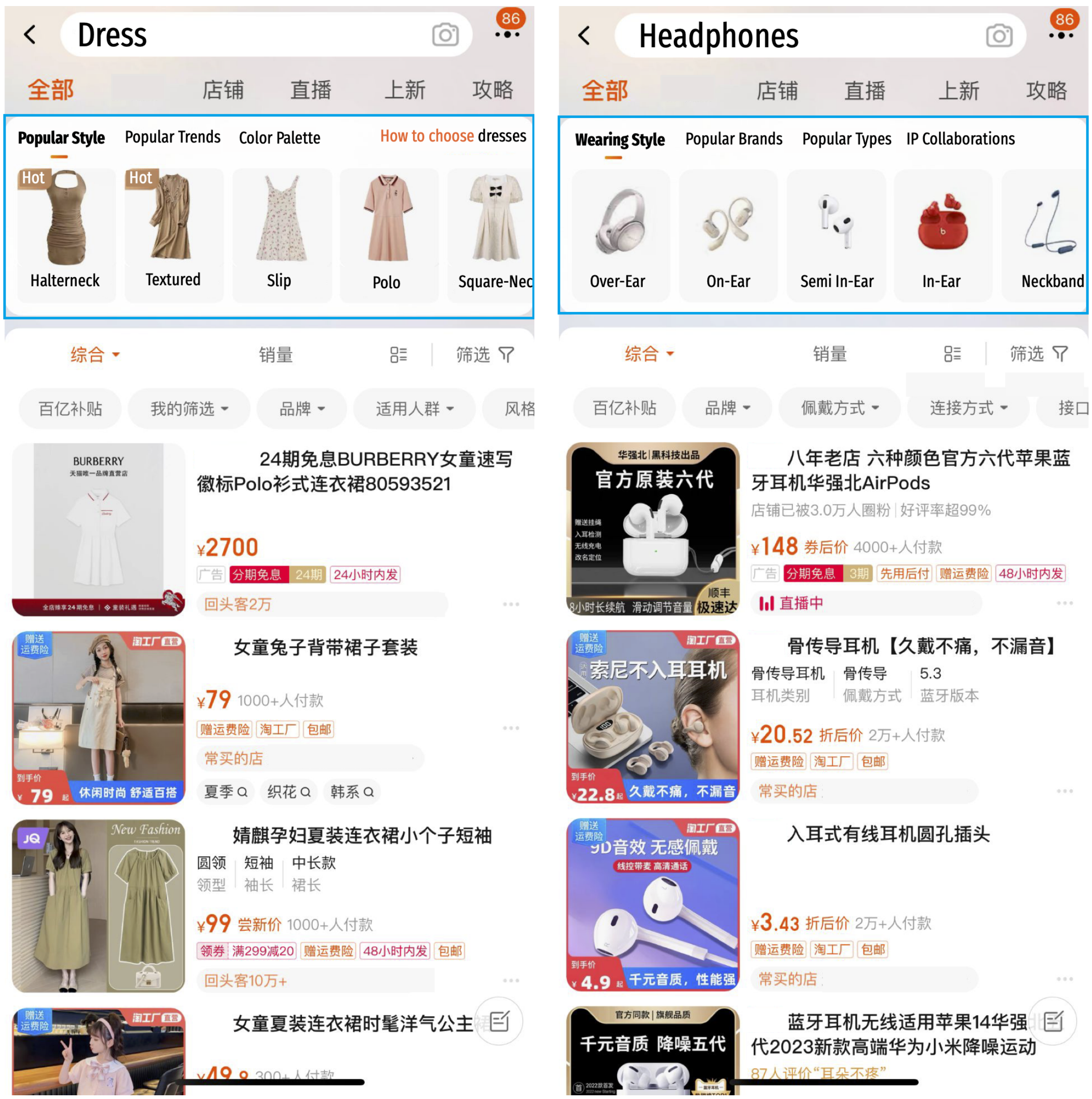
Customers can click on any of the PTG elements to refine their search. When they click on one of those elements, the search engine will automatically refine the search query to reflect the finer subset of relevant products. For example, if the customer clicks on the picture for the halterneck dress, 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."

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 Racquet."

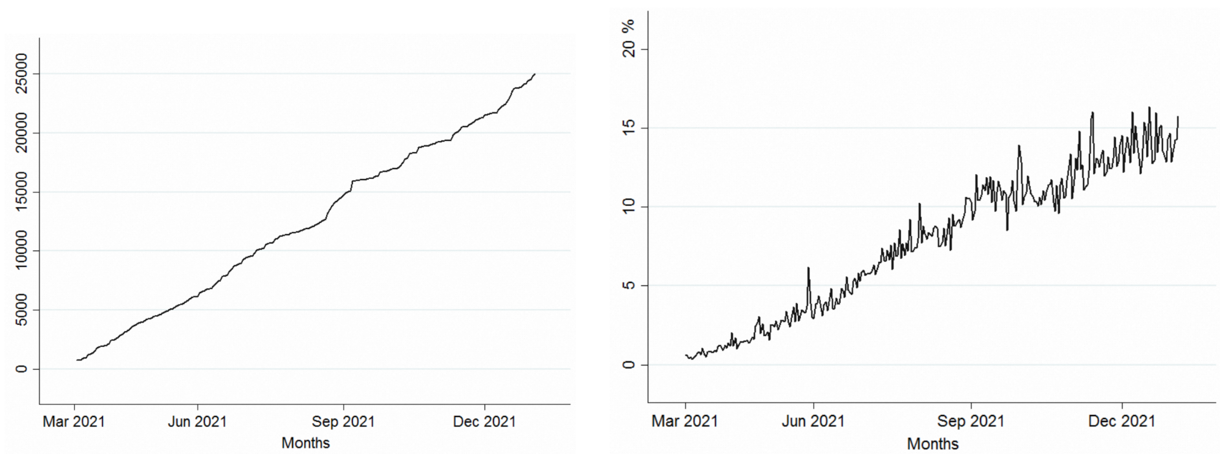


(a) Dress

(b) Headphones

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

The platform launched PTG in March 2021 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 general queries went from representing 0 to 15% of the gross merchandise volume (GMV henceforth) directly associated to a search query.



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

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

Figure 2: Expansion of the Picture-Text Guidance Search Tool 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 platforms' customers (existing and new) were ex-ante randomly allocated to a control and a treatment groups with equal probability.[4] 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. 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

[4]Consumers who were not logged in when searching for products faced are not included in our experiment.

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. 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. Additionally, the experiment allows us to test for heterogeneous effects of the search tool across customer cohorts that experienced progressively increasing access to the new search tool.

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);[5] 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

[5]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.

allows us to distinguish between more and less popular products or sellers, and how PTG affects consumer choices for the two 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, 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).[6]

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 6 months until the end of 2021.

4 Empirical Approach and Results

We evaluate the effect of PT guidance on customer search behavior and purchase decisions. To do so, we conduct our analysis at the individual customer 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. $Treatment_i$ is an indicator of whether the customer belongs to the treatment group, so the coefficient β measured the causal effect of interest.

We estimate the regression for several outcomes y_i. We focus on the number of products each customer views, the number of clicks they make, the total orders they place, and the overall sales value (GMV) they generate. These metrics are compiled at the individual consumer level for a designated period. Given recent concerns around using log transformations when outcomes can take the value zero (Chen and Roth, 2023), we estimate regressions in

[6]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.

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		3.8480	3.8518	0.4463
		0.0036	0.0036	
Number of Registered Years		6.2573	6.2636	0.5044
		0.0067	0.0066	
Female		0.5455	0.5454	0.9707
		0.0010	0.001	
View in the Past 8 Weeks		2585	2563.9	0.0865
		8.95	8.8	
Clicks in the Past 8 Weeks		112	111.55	0.3835
		0.39	0.39	
Orders in the Past 8 Weeks		4.53	4.55	0.4997
		0.02	0.02	
GMV in the Past 8 Weeks		410.6	415.43	0.4888
		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.

levels.

We present results separately for the short-run and long-run. In the short-run, we aggregate the outcomes of interest over the course of the first 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).[7]

4.1 Short-Run Results

We first show that the PT guidance tool had an immediate impact on altering customers' search behavior. To show this we run regressions in equation 1 for three separate outcomes: number of searches, number of PTG general searches, and number of PTG specific searches performed on the day of entry in the experiment. In the example provided in the introduction, the term "Dress" represents a general PTG query, whereas "Halterneck Dress" represents a PTG specific query. Table 2 displays the results. First, we find that treatment group customers search 0.049 (1.04%) more queries compared to the control group. To explore the reasons behind the increase in searches, we further examine customer searches associated with PTG queries. We find that customers search for more PTG specific queries, whereas their searches for general PTG remain unchanged. Treatment group customers search 0.053 (500%) more PTG specific queries compared to the control group. Therefore, the rise in the overall number of searches among the treatment group customers is primarily attributed to their increased searches for PTG specific queries. The results suggest that treatment group customers are actively utilizing the newly introduced PTG feature, and they adopt the PTG specific queries in their searches.

Next, we investigate if the introduction of PTG influenced the purchasing decisions of customers. Table 3 shows the treatment effects on product views, clicks, orders, and GMV each customer generates on the day of customer entry in the experiment. All coefficients in columns 1 to 4 are statistically insignificant from zero, indicating that the launch of PTG does not immediately influence the number of products customers observe, click on, or purchase.

[7] This guarantees that we can track all customers in our sample for 24 weeks.

Table 2: Short-Run Impact on Number of Searches

	All	PTG General	PTG Specific
	(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 according to equation 1.

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%
Constant	399.0***	17.30***	0.709***	50.14***
	(8.748)	(0.345)	(0.0262)	(5.407)
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 according to equation 1.

While there are not significant changes in the purchasing decisions of the treatment group customers concerning views, clicks, and orders, we further investigate if customers shift their searches from PTG general queries to PTG specific queries. Specifically, we analyze the effects of the treatment on purchasing decisions related to PTG specific and PTG general queries, respectively. Table 4 shows the results. For the treatment group customers, the results show a significant decrease in the total views, clicks, and sales stemming from PTG general queries. Conversely, these metrics saw a substantial increase for PTG specific queries. In terms of scale, the treatment group customers experienced over 500% increase in the total number of views, clicks, and sales for PTG specific queries compared to those in the control group. These results suggest that customers in the treatment group, in the intensive margin, are shifting their searches from PTG general queries towards PTG specific queries.

Table 4: Decompose Treatment Effects for PTG Query Words

	PTG General				PTG Specific			
	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 according to equation 1.

Overall, the results suggest that, after the implementation of PTG there is notable alteration in the search behavior of the treatment group. Consumers in the treatment group increase their number of searches, particularly for PTG specific queries, and shifting away from PTG general queries. Yet, apart from the increase in number of searches, there were no significant increase in the number of product views, clicks, or overall sales for the treatment group. Thus, while the PTG impacts customer search behaviors, it does not influence their purchasing decisions in the short run.

The lack of immediate influence from PTG on purchasing decisions might be attributed to customers requiring time to familiarize themselves with the product features of the newly discovered PTG specific queries and to experiment with these new products. Consequently, relying solely on the short-run results might not provide an accurate estimation of the search tool's positive benefits. As a result, we proceed to study the long-run outcomes of the treatment by monitoring customers for 24 weeks post the PTG implementation.

4.2 Long-Run Results

To investigate the long-run effects of introducing PT guidance, we monitor both the treatment and control groups for six months to analyze the influence on their search behaviors and purchasing decisions.[8] Given that PTG did not influence customer purchasing decisions in the short run, we first explore whether the purchasing decisions of those in the treatment group change over the long run. We run regressions in equation 1 by aggregating the outcomes of interest over the course of 24 weeks following a customer entry in the experiment, and examine the impact of PTG on product views, clicks, and orders generated. Table 5 shows the main long-run treatment effects. In contrast to the short-run results, customers in the treatment group show a significant increase in product orders (1.57%) and purchases (3.24%) over the long run. On the other hand, we do not find significant difference between the treatment and control group customers in terms of product views or clicks over the long run. These findings suggest that the use of PT guidance leads to increased final sales in the long run. This increase seems primarily due to the treatment group customers' searches becoming more effective, rather than them dedicating more time to viewing and clicking on more products.

Table 5: Long-Run Treatment Effects

	Views	Clicks	Orders	GMV
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 according to equation 1.

Given the different effects on total sales between the short and long run, we proceed to examine how these effects change over time. Specifically, we replicate the regressions from equation 1 by narrowing down to distinct time intervals following the commencement of the experiment. The first two columns of Table 6 report the results when we restrict the customers purchase order and GMV to the first week. Columns (3) and (4) report customer purchase in the first two weeks while Columns (5) and (6) report customer purchase in the first three weeks. Columns (7) and (8) report customer purchase in the first four weeks. Results in Table 6 indicate that, firstly, the magnitude of the coefficients for both

[8] Customer long-run behavior does not include their initial day of PTG search. Therefore, long-run and short-run are mutually exclusive. In the long-run analysis, we focus on customers who search the PTG queries for the first time before July 16 to ensure that we can observe all customers in the sample for 24 weeks, which includes $346,110$ customers.

order and GMV consistently rise as we transition from including only the first week's data to incorporating data from the first four weeks. Secondly, the treatment effects achieve significance (at the 5% level) when we use data from the first four weeks. The results confirm our earlier finding that 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.

Table 6: Long-Run Treatment Effects Over Different Durations

	Week 1		Week 1-2		Week 1-3		Week 1-4	
	Order	GMV	Order	GMV	Order	GMV	Order	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. % Change is calculated by dividing the treatment effect by the control group average. Standard errors are in parentheses. We include cohort fixed effects.

Next, we delve deeper into who get affected more by the implementation of PT guidance. We begin by examining the treatment effects among different customer categories. We study the customer heterogeneity across five dimensions: Age tier, a ranking where a higher tier indicates older customers. City tier, a measure of the customer's residential location, with a higher tier representing smaller cities. Register year, which is calculated as the number of years a customer has been on the platform. Total transaction amount, which is calculated as the sum a customer transacted in the 8 weeks preceding the experiment, serving as an indicator of their likelihood and frequency to transact on the platform. Lastly, we also investigate if there are differential effects between female and male customers. Table 7 shows the results. We find no significant heterogeneous treatment effects between male and female customers, across different age tiers, city tiers, or based on their registration dates on the platform. However, we do find that customers who transact more 8 weeks prior to the experiment benefit more from the PT guidance. Customers who frequently transact on the platform are likely more familiar with its functionalities. As they spend more time on the platform daily, they are also probably more inclined to invest time in exploring the new features of PT guidance. This, in turn, could enhance the effectiveness of their searches following the introduction of PTG.

Besides customers, we then investigate which products and sellers benefit more from

Table 7: Long-Run Heterogeneous Treatment Effects (Customers)

	GMV	GMV	GMV	GMV	GMV
	(1)	(2)	(3)	(4)	(5)
Treat	0.63	63.50	-0.0151	59.93**	-212.1***
	(44.48)	(45.13)	(44.52)	(29.40)	(19.66)
Treat*Age Tier	21.01				
	(16.59)				
Treat*City Tier		-3.204			
		(10.74)			
Treat*Register Year			7.557		
			(5.895)		
Treat*Female				-15.21	
				(39.14)	
Treat*GMV in the past 8 weeks					0.474***
					(0.00696)
Constant	774.4***	755.3***	785.1***	755.3***	848.7***
	(126.6)	(126.9)	(127.6)	(126.6)	(125.2)
Observations	346,110	346,110	346,110	346,110	346,110
R-squared	0.508	0.508	0.508	0.508	0.515

the implementation of PT guidance. We first classify products according to their total sales (measured in GMV). The top 10 products has the highest sales among all the available products. Panel I of Table 8 shows the results. We find that products ranked 1,000 and beyond exhibit a substantial increase in the sales over the long run. In contrast, the top-ranking products show no significant changes in the long run. This discrepancy could be attributed to consumers' existing familiarity or exposure to the top-ranking products, resulting in minimal impact from PT guidance. On the other hand, tail products typically receive less attention from consumers, leading to more significant effects when PT guidance is introduced. When comparing the effects between products ranked beyond 10,000 and products ranked between 100 and 10,000, we can see that products ranked beyond 10,000 have a larger treatment effect. The findings indicate that the tail products appear to gain more advantages from the introduction of PT guidance.

We conduct a similar analysis for sellers. We categorize sellers based on their total GMV quantiles. Panel II of Table 8 shows the results. The findings align closely with what we observed for the products. The top sellers show no significant increase in the sales over the long run. In contrast, the Med-low and Low 20% sellers show a significant increase in the sales over the long run. Regarding scale, sellers ranked in the medium-low category experience the highest long-run sales boost, showing a 5% increase in comparison to the control group of sellers.

Table 8: Long-Run Heterogeneous Treatment Effects (Products and Sellers)

	Panel I: Product				
	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: Seller				
	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 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.

4.3 Consumer Learning

Next, we delve into the increase in sales over the long term, examining the long-run effects for both PTG queries and non-PTG queries. Table 9 shows that the sales from PTG queries for the treatment group customers are 5.34% higher than those for the control group customers. More interestingly, the results indicate that the treatment effects on sales not only influence PTG queries but also extend to non-PTG queries in the long term. Column 4 of Table 9 shows that sales from non-PTG queries for the treatment group customers are significantly higher (by 3.07%) compared to those of the control group customers.

Table 9: Long-Run 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 according to equation 1.

Based on these findings, we intend to explore the underlying reasons for the sales increase from non-PTG queries. Our hypothesis is that customers can enhance their demand formation and learn to express their needs more effectively in the search engine with the assistance of PTG. Customers might adopt words they learned from PTG specific queries and integrate them into their searches for non-PTG queries. For instance, a consumer could discover the term "Halterneck Dress" from a PTG-specific query and then utilize similar phrasing in other searches, like "Halterneck Top." These matched queries, in turn, could contribute to a higher increase in orders and final sales for non-PTG queries.

To verify the hypothesis, We begin by analyzing how PT guidance affects the query length used by customers. Table 10 shows the long-run treatment effect on the average query length. We find that treatment group customers' search queries are 0.31% longer compared to the control group customers in the long term (as shown in column 1). This trend of using longer query words is evident not just in PTG-specific searches (column 3) but also in non-PTG searches (column 4). The findings indicate that following the introduction of PT guidance, customers of the treatment group have become more proactive in exploring new queries, particularly longer ones. Additionally, they are providing more detailed information during their searches using these longer queries.

Table 10: Mechanism of Customer Learning (Avg Query Length)

	All Queries	PTG General	PTG Specific	Non-PTG
	(1)	(2)	(3)	(4)
Treat	0.0191***	-0.000470	0.132***	0.0126**
	(0.00479)	(0.00339)	(0.00739)	(0.00535)
% Change	0.31%	-0.01%	1.80%	0.20%
Observations	346,110	346,110	346,110	346,110
R-squared	0.021	0.050	0.939	0.045

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.

To further investigate the mechanism, we analyze the types of words customers include in these longer queries. We categorize non-PTG queries into two types: non-PTG matched queries and non-PTG unmatched queries. Non-PTG matched queries refer to queries that share some similarities with a PTG specific query. For instance, if "Halterneck Dress" is a PTG specific query, then "Halterneck Top" would be classified as a non-PTG matched query. Essentially, we aim to determine if customers adopt the usage of PTG specific queries even in non-PTG searches. We investigate the influence of PT guidance on the number of searches for various query types. The long-term results regarding the number of searches are presented in Table 11. Firstly, column 3 in Table 11 shows that the treatment group's customers search for PTG specific queries are significantly higher than the control group. This aligns with our short-run findings. The difference in magnitude is approximately 450% when compared to the control group, mirroring the short-run results as well. This suggests that the increase in the number of searches for PTG specific queries remains consistent over time. Secondly, column 4 of Table 11 shows that customers in the treatment group tend to search more for matched non-PTG queries. However, there is not a significant difference when it comes to searches for unmatched non-PTG queries. The results suggest that PT guidance does not simply motivate consumers to conduct more searches in general. Instead, consumers explore combinations of PTG-specific inspired words with non-PTG queries, leading to increased search activity in this particular context.

Table 11: Mechanism of Customer Learning (# of Searches)

	All Query	PTG General	PTG Specific	Non-PTG Matched	Non-PTG Unmatched
	(1)	(2)	(3)	(4)	(5)
Treat	0.172	-0.0212	0.670***	0.313**	-0.790
	(0.814)	(0.0583)	(0.00523)	(0.133)	(0.682)
% Change	0.08%	-0.14%	450.55%	2.41%	-0.40%
Observations	346,110	346,110	346,110	346,110	346,110
R-squared	0.116	0.034	0.048	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 to equation 1.

Lastly, we examine whether the matching quality between consumers and products is improved in the long-run. Table 12 shows the results on rate of positive ratings and refund rate. The results show that customers in the treatment group demonstrate higher positive ratings and a reduced likelihood of refunds in the long run. These results suggest that the implementation of PT guidance indeed improves the match between consumers and products over time.

Table 12: Long-run Review Comparison

	Positive Rating Rate	Refund Rate
	(1)	(2)
Treat	0.00636***	-0.00274***
	(0.000201)	(0.0000731)
% Change	17.88%	-8.52%
Observations	7,479,300	7,479,300
R-squared	0.003	0.012

Notes: The dependent variables in the order level analysis equals to 1 for positive ratings/refund and 0 for negative ratings/non refund. Standard errors are in parentheses. We include cohort fixed effects according to equation 1.

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 their preferences. Leveraging a long-run experiment linked to the launch of PT guidance, we find that having access to textual and visual search recommendations increases purchases by 3.24%. The increase does not seem to be driven by consumers viewing or clicking on more products, suggesting an increase in search effectiveness. The increase also is not only driven by searches that are directly affected by PT guidance, but rather it spills over to other searches, implying that the tool is capable of teaching customers how to perform better searches on their own.

Although the experiment directly affected demand, we find important indirect effects for sellers. In particular, the ability of PT guidance to narrow down searches to subsets of more specific categories benefit products and sellers outside of the most popular.

Importantly, we also find that if we restrict our analysis to the short-run, we would not be able to detect the significant benefits of the new search tool. This null result manifests itself in two dimensions. First, we find that the immediate effect of the new search tool on the first day a customer enters the experiment is a precisely estimated zero. This result likely reflects the selection of people who enter the experiment as individuals who are visiting the e-commerce platform for the explicit purpose to buy (or not buy) something. It also implies that it takes time for search tools like PT guidance to display their beneficial effects on consumers. Second, using only data from the first month of the experiment, the estimated effect, although positive and comparable to the long-run estimate, is too noisy to be distinguished from zero in a short time span.

Our results have a number of important implications, both for the design of search and matching mechanisms and for the design of experiments. On search and matching, our results highlight the importance of addressing demand expression and formation challenges. In this paper, we evaluated the possibility of adding a combination of textual and visual cues. An important avenue for future exploration would be to separate the role of textual versus visual suggestions.

On the design of experiments, our results highlight the risk of drawing conclusions from short-run results. In our case, if we had only had access to one week or one month of data, we would have concluded that the search told was not effective in helping consumers identify their preferences because the estimates are too noisy to detect a significant effect. It is important to expand research to understand how to draw long-term conclusions from short-term experiments.

The paper has a number of limitations. First, we are unable to evaluate whether the search tool, by recommending searches that the consumer would not have otherwise performed, diverts 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, we find that two proxies for customer satisfaction – ratings and returns – are affected and in a direction that suggest consumers are more satisfied with the purchases they make in the treatment group compared to the control group. 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 for products that they would not otherwise want to buy.

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.

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

Chen, Yuxin, Zhe Yuan, Tianshu Sun, and AJ Chen, "Understanding the Impacts of De-personalization 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.

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.

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.

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.

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.

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.

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.

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

Appendix

A More Details about Picture-Text Guidance

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" (as shown in the upper channel (1) of Figure A.1), or click on the picture with text "Oat" and search for products related to "Dress Oat" (PTG specific query) instead of just "Dress" (as shown in the lower channel (2) of Figure A.1).

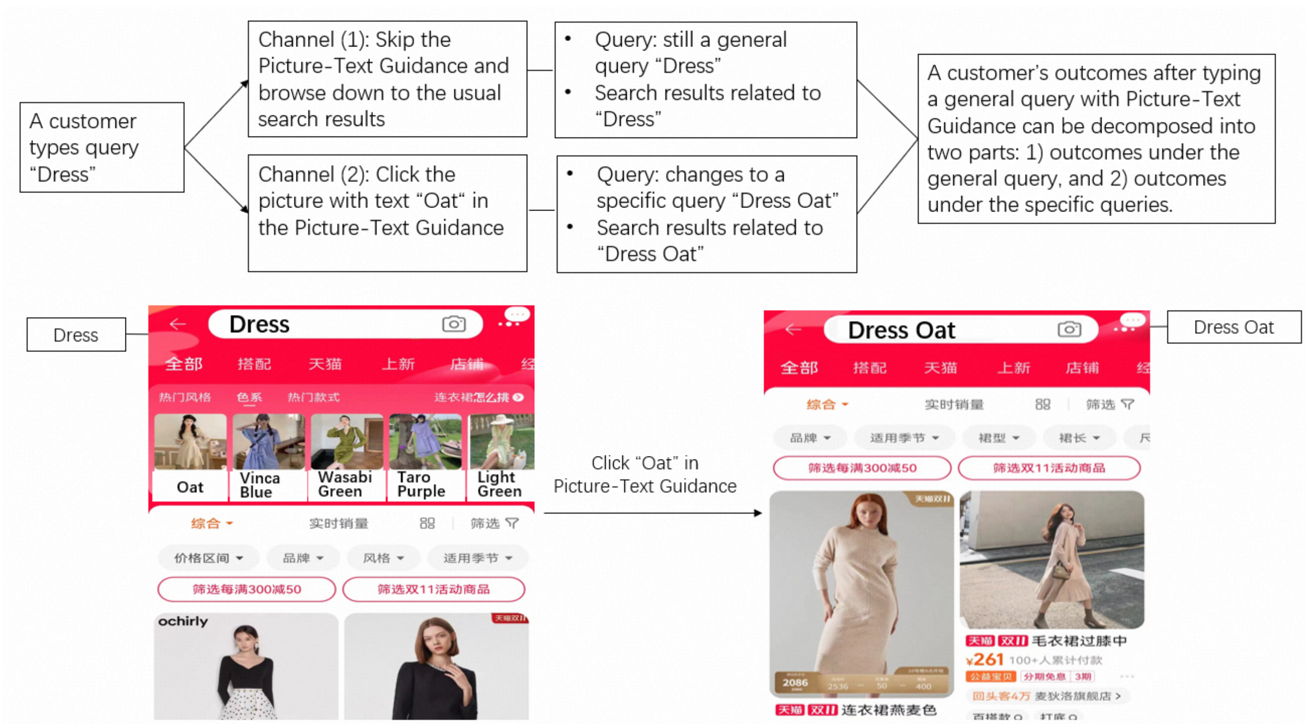


Figure A.1: Customer Search Process with Picture-Text Guidance

B Robustness Check

Table B.1 is identical to Table 5 except that we control for customer characteristics. The main long-run treatment effects results exhibit a high degree of similarity in terms of both sign and significance level with Table 5. Customers in the treatment group show a significant increase in product orders (1.47%) and purchases (2.66%) over the long run. On the other hand, we do not find significant difference between the treatment and control group customers in terms of product views or clicks over the long run. These findings suggest that the use of PT guidance leads to increased final sales in the long run.

Table B.1: Long-run Treatment Effects (Controlling for Customer Characteristics)

	Views	Clicks	Orders	GMV
Treat	-29.00	-0.719	0.316**	51.34***
	(52.05)	(2.130)	(0.126)	(19.41)
% Change	-0.23%	-0.14%	1.47%	2.66%
Constant	14,297***	629.4***	21.02***	759.7***
	(338.1)	(13.84)	(0.817)	(126.1)
Observations	346,110	346,110	346,110	346,110
R-squared	0.130	0.15	0.218	0.508

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 B.2 presents the long-run quantile regression results for orders. We observe that only customers in the 90th quantile (customers with a higher number of orders) exhibit an increase in purchases. Table B.3 reports the long run quantile regression results for GMV. We find that 25% quantile and 90% quantile customers purchase more. This indicates that the effects of the PTG may not be uniform across all customers. Instead, it appears that the most active customers take advantage of the PTG and subsequently make more purchases, contributing to the overall average increment in purchase orders.

Table B.2: Long-run Quantile Regression

	10%	25%	50%	75%	90%
Treat	0	0	0	0	1**
	(omit)	(omit)	(omit)	(omit)	(0.470)
Observations	346,110	346,110	346,110	346,110	346,110

Table B.3: Long-run Quantile Regression

	10%	25%	50%	75%	90%
Treat	0	3.740***	1.750	10.17	74.84**
	(omit)	(0.978)	(3.624)	(8.428)	(29.72)
% Change		3.04%	0.35%	0.74%	2.21%
Constant		344.9****	937.2***	2.374***	5.572***
		(19.62)	(35.57)	(81.56)	(231.1)
Observations	346,110	346,110	346,110	346,110	346,110

C Extra Results for the Analysis

C.1 Additional Short-Run Results

For the short-run analysis, we focus on the first day of PTG search, and evaluate the impact of introducing Picture-Text Guidance on the treatment group. The results suggest that for PTG queries, the treatment group consumers show higher product views (2.24%). We do not find significant changes for queries that are not PTG.

Table C.1: Short-Run Treatment Effects

	PTG Query				Non-PTG Query			
	Views	Clicks	Orders	GMV	Views	Clicks	Orders	GMV
Treat	1.354***	0.0212*	0.00371***	-0.157	-0.687	0.0352	0.00166	0.507
	(0.284)	(0.0124)	(0.00119)	(0.27)	(1.081)	(0.0427)	(0.0031)	(0.638)
% Change	2.24%	0.93%	2.97%	-1.78%	-0.35%	0.47%	0.53%	2.14%
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	0.018	0.029	0.01	0.002

We examine whether PTG leads to improved matching between consumers and products. To assess the quality of matching, we focus on two measures: the positive rating rate and the refund rate of orders. These are the short-run results analog of the long-run results in Table 12.

Table C.2: Review and Refund across Treatment and Control Groups (Short-run)

	Positive Rating Rate	Refund Rate
Treat	-0.00391	-0.000633
	(0.00275)	(0.00191)
% Change	15.22%	7.10%
Observations	64,074	64,074
R-squared	0.007	0.009

Notes: The dependent variables in the order level analysis equals to 1 for positive ratings/refund and 0 for negative ratings/non refund. Standard errors are in parentheses. We include cohort fixed effects according to equation 1.

Table C.2 reports the rate of positive ratings (column 1) and refund rate (column 2). The Results suggest that the impacts on the positive rating rate and refund rate are insignificant.

C.2 Additional Long-Run Results

Table C.3 depicts long-run review under PTG queries.The results are consistent with those in Table 12, which reports the long-run review for all queries. The results indicate that customers in the treatment group exhibit higher positive ratings and a decreased probability of requesting refunds over an extended period for PTG queries. These findings strongly suggest that the implementation of PT guidance effectively enhances the alignment between consumers and products as time progresses.

Table C.3: Long-run Review Comparison for PG queries

	Positive Rating Rate	Refund Rate
	(1)	(2)
Treat	0.00515***	-0.00363***
	(0.000686)	(0.000384)
% Change	15.70%	7.93%
Constant	0.161***	0.0763***
	(0.00379)	(0.00212)
Observations	713,883	713,883
R-squared	0.002	0.004

Notes: The dependent variables in the order level analysis equals to 1 for positive ratings/refund and 0 for negative ratings/non refund. Standard errors are in parentheses. We include cohort fixed effects according to equation 1.

Table C.4 presents the results for long-run heterogeneous treatment effects in Orders, serving as analogs to the results shown in Table 8. Notably, the analysis reveals that the PTG confers advantages to tail products and small sellers when compared to top-ranking products and top sellers. This finding highlights the positive impact of the PTG on promoting visibility and sales for less prominent products and sellers within the marketplace.

Table C.4: Long-Run Heterogeneous Treatment Effects (Order)

	Panel I: Product				
	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: Seller				
	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 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.