Marketing Science Institute Working Paper Series 2021

Report No. 21-114

The Effect of Voice AI on Consumer Purchase and Search Behavior

Chenshuo Sun, Zijun (June) Shi, Xiao Liu,

Anindya Ghose, Xueying Li, & Feiyu Xiong

The Effect of Voice AI on Consumer Purchase and Search Behavior

Chenshuo Sun
Stern School of Business
New York University
csun@stern.nyu.edu

Zijun (June) Shi
Business School
HKUST
juneshi@ust.hk

Xiao Liu
Stern School of Business
New York University
xliu@stern.nyu.edu

Anindya Ghose
Stern School of Business
New York University
aghose@stern.nyu.edu

Xueying Li
Alibaba Inc.
xueying.li@alibaba-inc.com

Feiyu Xiong
Alibaba Inc.
feiyu.xfy@alibaba-inc.com

Abstract

Voice-activated shopping devices (voice AI), such as Amazon's Alexa or Alibaba's Tmall Genie, as a new channel for shopping, are gaining popularity among consumers worldwide. It has become important, therefore, to understand how the voice AI adoption affects consumer shopping behaviors. In this paper, we use large-scale archival data containing consumer-level browsing and purchase records from one of the world's largest online retailers, Alibaba, who has developed and launched its own voice AI product called Tmall Genie to empirically identify our results. The results show that consumers on average spend 23% more as a result of adopting the voice AI, corresponding to approximately US$ 630 million increase in sales revenue every year. The effect on purchase is stronger for younger consumers and male, and the effect on browsing is more salient for younger consumers and female. We also find that although the positive effect on purchases shrink over time, it still remains significant after eight weeks; whereas the effect on browsing is only significant immediately after the adoption. Moreover, the voice channel does not cannibalize other purchase channels (PC and mobile); rather, it has a positive spillover effect on the mobile channel. Lastly, the impact of voice AI adoption on purchase is stronger for products that do not require active search or comparison, such as product categories with low substitutability or high purchase frequency. To summarize, we demonstrate that the adoption of voice AI has enhanced the growth of the digital sales market in the world's largest e-commerce platform. The findings provide useful implications for both e-commerce companies and businesses that harness voice-assisted shopping.

Keywords

Voice assistant ▪ voice shopping ▪ e-commerce ▪ causal impact ▪ heterogeneity ▪ channels

1. Introduction

Voice-controlled smart devices are gaining popularity. In 2020, there will be 4.2 billion digital voice assistants being used in devices around the world.[1] In the U.S., households now own roughly 118.5 million devices[2] such as Amazon Alexa (and Echo), Google Home, and Apple Homepod (Figure 1 upper panel). In China, millions of consumers have adopted Tmall Genie, Baidu Xiaodu, and Xiaomi Xiaoice (Figure 1 bottom panel). [3]

Figure 1. Examples of Voice-controlled Smart Devices

Amazon Echo	Google Home	Apple Homepod
		
Alibaba Tmall Genix	Xiaomi Xiaoice	Baidu Xiaodu
		

As these devices take off, voice-activated shopping—or placing orders by talking to artificial intelligent (AI) devices—is showing signs of considerable growth. For example, Amazon Alexa, launched in 2014, saw voice-shopping triple during the 2018 holidays versus the year earlier. In April 2019, Walmart collaborated with Google to bring "Walmart Voice Order" which allows consumers using Google Home to purchase products directly from Walmart's online shopping portal by simply telling the voice AI what they want (Cohen, 2019). A survey by Adobe Analytics found that 22 percent of home speaker owners polled use their devices for shopping,[4] and another

[1] Cited from https://www.statista.com/statistics/973815/worldwide-digital-voice-assistant-in-use/
[2] Cited from https://www.cbinsights.com/research/facebook-amazon-microsoft-google-apple-voice/?utm_source=CPG+%26+Retail+Insights&utm_campaign=b0d7d8c243-CPGNL_04_24_19_copy_50&utm_medium=email&utm_term=0_5a34af6e3b-b0d7d8c243-91840793
[3] The contemporary use of the major voice assistant devices is provided in the Appendix.
[4] Cited from https://www.marketwatch.com/story/the-booming-smart-speaker-market-and-the-services-it-will-help-2018-01-18

study by PwC found that 10 percent of the U.S. voice-AI users it sampled place orders by voice on a daily basis.[5]

Despite the significance of voice shopping to our daily lives and the global economy, little is known about how consumers' shopping behavior is altered by voice-activated shopping. Voice shopping can affect a consumer's purchase journey, from consideration (search) to the final transaction. For example, a consumer who sees a video featuring lipsticks by a popular cosmetics blogger can immediately ask, "Alexa, what are the best-selling lipsticks?" to start the search process and finally place an order by saying "Checkout my cart."[6] The direction of the impact on purchase or search is a priori unclear. Two contradictory effects could be at work simultaneously. On the one hand, a voice AI allows multi-tasking and enables us to stay hands-free, thereby making it easier for consumers to find a product and complete a transaction, leading to smoother searches and increased purchases. On the other hand, research on auditory consumer behavior indicates that information presented by voice is usually more difficult to process than that presented in text or visually (e.g., Munz and Morwitz, 2020), which may lead consumers to defer their searches and purchases. Moreover, a digital device (like a tablet or a smartphone) or a brick-and-mortar store, compared to a voice shopping channel, displays a far broader range of products and facilitates impulse shopping (Baumeister, 2002; Zhang and Shrum, 2009). Therefore, an empirical study is essential to help provide a better understanding of whether shopping by voice will resonate with consumers, guiding managers toward improving consumer experience.

This research takes the first step to study empirically how consumers' purchase and search behavior changes due to the adoption and usage of voice-activated shopping in an e-commerce

---

[5] Cited from https://www.pwc.com/us/en/services/consulting/library/consumer-intelligence-series/voice-assistants.html?utm_campaign=sbpwc&utm_medium=site&utm_source=articletext

[6] Cited from https://www.racked.com/video/2018/2/20/17031958/alexa-shopping-amazon-echo-dot-voice

setting. Specifically, we ask five questions. First, how does the voice AI device adoption change consumers' purchase and browsing behavior? Second, how does voice AI device adoption heterogeneously affect different groups of consumers? Third, what are the changes in purchase and browsing behaviors in the short-run and long-run? Fourth, does the voice channel cannibalize other traditional purchase channels (PC and mobile)? And fifth, how does voice AI adoption affects consumers' shopping behavior for different categories of products?

To examine these questions, we use large-scale archival data from one of the world's largest online retailers, Alibaba, who has developed and launched its own voice AI product, Tmall Genie, to provide a voice shopping experience for consumers, similar to that of Amazon Alexa. Our unique transactional dataset contains consumer-level browsing and purchase records for both voice AI adopters and non-adopters for a 31-week period. We employ propensity score matching to match adopters with non-adopters based on their observable characteristics such as demographics and shopping behavior prior to the adoption. We then use the difference in difference approach that controls for unobserved user-level and time-varying effects to identify the effects of adopting a voice assistant on purchase and browsing behavior in the post-treatment period. Our results yield several important findings. Overall, we find that each consumer spends 23% more on average, as the result of adopting the voice AI product, corresponding to an approximately RMB 4.4 billion (US$ 630 million) increase in annual sales revenue every year. Second, we find that there is a substantial heterogeneity in the treatment effects. The effect on purchase is more salient for younger consumers and male, and the effect on browsing is more salient for younger consumers and female. Third, we find that although the positive effects shrink over time, they still remain statistically significant after 8 weeks. Fourth, the voice channel does not cannibalize other purchase channels (such as the PC channel and the mobile channel); rather, it has a positive spillover effect

on the mobile channel. Lastly and importantly, we propose and test two hypotheses about how voice AI adoption affects consumers' shopping behavior for different types of products and show that the mechanism has to do with search cost reduction. Specifically, the impact of voice AI adoption on purchase is stronger for products that do not require active search or comparison, such as product categories with low substitutability or high purchase frequency. To summarize, we demonstrate that the adoption of voice AI has enhanced the growth of the digital sales market in the world's largest e-commerce platform.

This paper takes the first step towards understanding the role of voice-activated shopping in changing consumers' shopping behavior. Our findings have direct managerial implications for e-commerce companies. Voice activated devices have been taking care of several tasks for customers—playing music, providing real-time information, making calls, setting alarms, etc. What sets apart e-commerce's voice-activated shopping, especially in a world where IT giants like Apple and Google are major players, is its rich connection with a customer's purchase history on the platform and the ability to allow shopping across large marketplaces by voice. To further leverage such advantages, a comprehensive understanding of how consumers react to voice-activated shopping is important. According to our findings on heterogeneous treatment effect of the voice AI adoption, managers may want to promote the voice AI device to consumers who tend to spend, purchase, or search more after using the smart device. Our findings on category-level analysis suggest that managers may advertise the voice AI device together with the low substitutability and high purchase frequency categories. Moreover, our research may also provide guidance for businesses beyond e-commerce, including the hospitality industry. For example, Marriott hotels now include Amazon Alexa; guests can order room service by talking to the voice

activated ordering devices in their rooms.[7] The findings on consumers' heterogeneous reactions to voice shopping may transfer to offline voice-aided purchase (or demand for services) in the context of hotels.

2. Relevant Literature

Our work builds on and extends several streams of research. First, this work is related to the growing body of literature on how voice assistants affect consumers and service providers. Voice assistant, also called smart speaker, refers to conversational agents that perform tasks with or for an individual (Mari et al. 2020). The act of placing orders online using a voice assistant goes under the name of "voice shopping" or "voice commerce". Major tech companies nowadays are increasingly commercializing voice assistant products and developing their shopping capabilities, which reveals voice assistant's disruptive potential for marketing (Dawa and Bendle 2018). Although practitioners are confronted with the rise of voice shopping, the study of how voice assistant, as a new shopping medium, affects consumption remains substantially uninvestigated. Existing literature is largely descriptive, theoretical, or experimental-based. Kumar et al. (2016) provide a framework to understand intelligent agent technology applications and their adopting using a grounded theory approach. Jones (2018) provides a case study to explore the implication, application, and opportunities for voice AI in marketing. Smith (2020) examines what type of marketing message is acceptable to consumers on voice AI via a survey. Ba et al. (2020) develop theoretical models of sales policies through a voice assistant. Munz and Morwitz (2020) conduct fifteen lab experiments and find that information presented by voice is more difficult for consumers to process than the same information presented in writing. Luo et al. (2019) exploit

[7] Cited from: https://www.reuters.com/article/us-amazon-com-marriott-intnl/amazons-alexa-will-now-butler-at-marriott-hotels-idUSKBN1JF16P

field experiment data to study how the disclosure of AI-assisted financial chatbot would affect customer purchase decisions. Luo et al. (2020) show that firms can benefit from hiring an AI coach over human managers when training sales agent. Despite these extant studies, the question of whether and how the adoption of a voice assistant affects consumption in a real-world setting remains unknown. In the present work, we tap into a fine-grained transactional dataset collected in the production environment of one of the most popular voice assistants, launched by one of the world's largest e-commerce platform, namely Alibaba Tmall Genie, to empirically quantify the effects of voice AI adoption on consumers' purchase and browsing behavior. Moreover, we estimate the heterogeneous treatment effects across user demographics and across product categories, thereby providing insights on how to better target users for voice shopping. To our best knowledge, our paper is the first to provide a comprehensive analysis on the impact of voice assistant adoption on consumers' purchase and browsing behaviors.

More broadly, our work also extends the line of literature that examines the impact of emerging technologies on consumer behavior and channel attribution. Ghose et al. (2013) explore how consumers' online browsing behavior varies between personal computers and mobile channels. They show that smaller screen sizes on mobile phones increase users' cost to browse for information. Wang et al. (2015) study how mobile devices affect consumer purchase behavior and show that purchase probability and the average order value increase as customers become accustomed to mobile shopping. Xu et al. (2014) examine consumers' news consumption behavior in response to the introduction of the mobile channel and find that the introduction of a mobile app leads to a significant increase in demand at the corresponding mobile news site. Xu et al. (2016) investigate the causal impact of tablets on e-commerce sales, and the complementary and substitution impact of the tablet channel on smartphone and PC channels. Ghose (2017)

summarizes how smartphones have reshaped consumer-buying behavior when consumers are exposed to mobile advertising by multiple forces. Xu et al. (2020) show that adopting new financial technologies such as mobile payment substantially increase consumer multichannel consumption behaviors. Tan and Netessine (2020) study the impact of tabletop technology on restaurant performance and find that the tabletop technology is likely to improve average sales per check and reduce the meal duration through improving service quality. What is missing in the literature, besides the main effect of voice assistant adoption on consumption, is how voice assistant as a new channel of shopping affects traditional purchase channels. The present study contributes to this strand of literature by looking into the spillover effect of voice assistant adoption on sales through PC and mobile channels.

3. Empirical Background and Data Description

We acquired access to a large archival data set of consumer-level purchase and browsing histories from the world's largest e-commerce platform, Alibaba. Specifically, the data come from its online business-to-consumer retail site, Tmall. Since its inception in April 2008, Tmall has acquired more than 726 million registered users. In 2019, Tmall's sales revenue exceeded RMB 5.7 trillion (US$ 853 billion),[8] accounting for 55.9% of online retail sales in China.[9] Traditional online shopping channels at Tmall include PC and mobile. In July 2017, Alibaba launched Tmall Genie, an AI-powered smart device to link customers with new shopping experiences, smart home appliances, and other new services. As with similar voice assistant products developed by other tech giants, with Tmall Genie,[10] consumers can use voice commands to conduct searches and make

[8] Cite from Alibaba: https://www.alibabagroup.com/en/news/press_pdf/p190515.pdf
[9] Cite from press: https://www.iimedia.cn/c1061/71838.html
[10] For the rest of the paper, we will use "Tmall Genie" and "Genie" interchangeably to refer to the Tmall Genie, which is the focal voice AI product.

purchases. More than 1 million orders were placed and paid via Genie's voice-shopping feature during Alibaba Group's "11.11" Global Shopping Festival in 2019.[11] Since the voice assistant AI has become a promising hit, Alibaba announced a plan to invest RMB 10 billion (US$ 1.4 billion) in the AI system for voice assistants in May 2020.[12] A typical conversation between a consumer and Genie in the voice shopping context begins with the customer's inquiry. In an illustrative example shown below, a consumer who wants to buy a mobile refill card has an efficient conversation with Genie and finally makes the purchase via the voice channel.[13]

Customer	"Tmall Genie, I'd like to order a mobile refill card."
Genie	"Master, I would recommend China Mobile's refill card. The total price is 100 RMB. It will be delivered to (address). May I place the order for you?"
Customer	"Yes, please place the order."
Genie	"Sure! In order to proceed, let us do voice authentication first. Please keep quiet around, and after the 'beep', say 'Tmall Genie, 2065.'" (Here 2065 is the authentication code randomly generated by the system.)
Customer	"2065".
Genie	"Alipay discount is applied. If you want to know the delivery status, you can let me know by saying 'Tmall Genie, tracking information.'"

Collaborating with Alibaba, we collected data of 39,885 randomly selected consumers from April 1, 2019 to October 31, 2019. Our data consists of 5.59 million consumer-level transactions, which encompasses 5.23 million purchases from mobile devices (including smartphones and tablets), 0.35 million purchases from PCs, and 0.005 million purchases from the voice shopping channel.[14] The purchase transaction data includes anonymized user ID, product category name, purchase date, confirmed payment amount, and the access channels (PC, mobile, and voice). Tmall has

[11] Cite from press: https://www.alizila.com/millions-in-china-tapped-voice-shopping-during-11-11-tmall-genie/
[12] Cite from press: https://economictimes.indiatimes.com/news/international/business/alibaba-to-invest-1-4-billion-in-ai-system-for-smart-speakers/articleshow/75842991.cms?from=mdr
[13] We provide two additional examples of using Genie for shopping in the Appendix.
[14] With voice-assisted purchases representing a small fraction of all purchases in our dataset, we note that estimates derived may apply more to the phenomenon of early penetration.

categorized all the products sold on the platform into 142 unique categories (see Appendix Table A3 for the full list). Our data also includes 132.41 million consumer-level browsing records in terms of unique pageviews. The browsing data includes anonymized user ID, product category name, visit date and count of pageviews, for the same set of consumers as included in the purchase data. These records indicate how many unique pageviews are made for each product category by each consumer on each day. Nevertheless, unlike the purchase data that has channel information, the browsing data does not have the channel information.[15] The dataset also comes with user demographics (i.e., *age* and *gender*) and the *total number of historical purchases* prior to the first day of data collection (April 1, 2019).

3.1. Sample of Genie Adopters and Non-adopters

Alibaba has reported that over 10 million Genie units have been activated, rendering Genie the most popular voice assistant product in China as of 2019.[16] With the above-described data, we leverage a small set of Genie adopters to conduct the analysis. In our case, "adopters" are those who adopted Genie and "non-adopters" are those who did not adopt the voice assistant. Our dataset has 710 adopters that are randomly drawn from the Genie adopters' pool, whose first-time adoption of Genie took place in the first week of July (namely, July 1 – 7), 2019.[17,18] The 39,175 non-adopters are randomly sampled from the entire pool of Tmall customers. These non-adopters did not adopt Genie any time before our data collection ended (on October 31, 2019). For the adopters, they can browse and make purchases on Tmall via all the three channels (namely, PC, mobile, and voice), whereas the purchase channels for the non-adopters are limited to PC and mobile.

---

[15] The pageview data includes three channels, PC, mobile, and voice. When a consumer asks Tmall Genie for information about a specific product, our data records it as one pageview via the voice channel.
[16] Cite from SEC: https://www.sec.gov/Archives/edgar/data/1577552/000110465920082881/a20-6321_46k.pdf
[17] Our data does not include those whose first-time adoption happened in other weeks.
[18] In our data, all adopters adopted Genie only once. We provide the model-free evidence showing the adoption rate by gender and by age groups in the Appendix Figure A1 and Table A1.

One potential concern with this setting is that consumers may choose to adopt Genie due to some shocks that are not random (for example, promotion). We confirmed with Alibaba that this is not the case, and that these adopters adopted Genie without any promotion activities going on during that week.

To identify the impact of Genie adoption on changes in online shopping behavior, a key challenge is to address the self-selection issue. For instance, it may be plausible that Genie adopters have greater disposable income in general and are likely to purchase more on Tmall over time. Consequently, consumers who belong to the adoption group may appear to purchase more on Tmall than the non-adopters. We employ several econometric techniques to account for endogeneity concerns, which we describe in detail in Section 4.

3.2. Measuring Purchase and Browse

We measure consumers' purchase and browsing behaviors using a set of metrics. The metric used to measure an individual's purchase behavior is the *total spending amount* (which is the confirmed payment amount across all 142 categories). Besides the overall spending on all categories, we calculate the total spending amount for each category, thereby producing category-level purchase metrics. We use the number of pageviews to operationalize an individual's browsing behavior. Similarly, we count the *total pageviews* on all categories as well as on each category. These metrics constitute the dependent variables of this study.

We exclude the purchase and pageviews of Genie itself from the calculation of the above metrics; otherwise, the outcome variables of the two groups would not be comparable in the first place due to the adoption of Tmall Genie. Because purchase and browsing at the user-day level could hugely fluctuate, we aggregate the raw data to the user-week level. Our data collection period (April 1, 2019 – October 31, 2019) corresponds to 31 weeks (i.e., from the week 14 through the week 44 of

2019), making it a panel dataset of 1,236,435 observations for 39,885 consumers. Among the 31 weeks, 13 weeks (April 1 - June 30, 2019) belong to the pre-treatment period, 1 week (July1 - July 7, 2019) is the treatment period, and 17 weeks (July 8 - October 31, 2019) belong to the post-treatment period. We report the summary statistics in Table 1.

Table 1. Summary Statistics[19]

Variable	N	Mean	SD	Min.	Max.
User Demographics					
Age	39,885	32.03	7.735	16	57
Gender (male = 1)	39,885	0.578	0.4942	0	1
Number of Historical Purchase	39,885	295.2	466.3	0	13,641
Purchase and Pageview					
Total Spending (RMB)	1,236,435	19.59	312.4	0	246,251
Total Pageviews	1,236,435	140.9	284.6	0	15,659

Note: Summary statistics are calculated on an unmatched sample of 710 adopters and 39,175 non-adopters observed over 31 weeks starting April 1, 2019 through October 31, 2019. The unit of analysis is an individual user for demographics and a user-week for the purchase and pageview variables.

4. Empirical Model

4.1. Identification Strategy

Our goal is to identify local average treatment effects (LATEs) of adopting Genie on consumer's purchase and pageview behaviors. The major challenge we face is that we do not have a randomized assignment of consumers into treatment and control groups. A widely adopted approach is to estimate a difference-in-difference (DiD) model, which eliminates persistent linear and additive individual specific effects that may introduce endogeneity into adopting Tmall Genie due to self-selection. On top of this, we assume that we can control for the unobserved need for purchase and browsing in absence of the voice-activated channel by conditioning on a rich set of observed characteristics (for a similar approach, see Datta et al. 2018; Bronnenberg et al. 2010);

[19] Throughout the paper, we use 4 significant figures (e.g., 1.000, 10.00, and 100.0) to report the results.

that is, we use a quasi-experimental matching procedure in which we match adopters with similar non-adopters based on a propensity score constructed from variables that capture users' shopping behavior and demographics.

4.2. Comparison of Adopters and Non-adopters

We first assess whether adopters and non-adopters display the same purchase and browsing behavior prior to adoption. We compute the measures of shopping behavior during the pre-adoption period (namely, from April 1 to June 30, 2019). Table 2(a) shows that, on an average, adopters and non-adopters differ significantly on some key demographic and behavioral characteristics. Adopters are younger than non-adopters (30.77 years old versus 32.06 years old, $p<0.001$). There is a lower percentage of male users in adopters than that in non-adopters (0.380 versus 0.582, $p<0.001$). In terms of purchase behavior, adopters' weekly spending amount is higher than non-adopters (RMB 35.93 versus RMB 19.36, $p<0.001$), and adopters have made more historical purchases (569.9 versus 290.2, $p<0.001$) than non-adopters. Adopters also conduct more pageviews than non-adopters (218.3 pageviews versus 139.4 pageviews, $p<0.001$). As we can see, consumers in the treatment and control groups differ significantly on several demographics and behavioral characteristics. It is likely that mere differences in the composition of the two groups could explain the difference in their shopping behavior. In other words, there exists self-selection into different groups. Therefore, we proceed to use quasi-experimental methods to deal with the selection issue.

4.3. Propensity Score Estimation and Matching

To reduce the potential difference between the treated and control groups thus alleviating the self-selection issue, we rely on matching strategies to get non-adopters who are similar to adopters on observable characteristics. Though individual fixed effects can partially account for the selection

issue, it imposes a linear functional form; whereas matching allows for more flexibility. As a result, each adopter is paired with non-adopters who have similar propensities of being treated. In this way, we can make a fair comparison of shopping behaviors between these two groups of consumers. In particular, we estimate each individual's adoption propensity as a function of observed variables:

$$\Pr(\text{Adoption}_i) = \Pr(\varphi_0 + \varphi Z_i + \epsilon_i > 0), \quad (1)$$

where Z_i is a vector of observed individual-specific characteristics. The covariates entering Z_i describe a user's demographic and purchase behavior during the 11 weeks prior to the adoption week (July 1-7, 2019), including age, gender, total spending, total pageviews, and number of historical purchases. We assume ϵ_i are independent and identically distributed random variables with a type I extreme value distribution, making the propensity score estimation a logit model.

Following a static matching estimation similar to that of Datta et al. (2018), we use the same pre-adoption observation window to predict consumers' adoption propensity. We then match each Genie adopter with non-adopters who resemble them most closely in terms of their overall propensity scores. In particular, we employ the nearest five neighbors matching with replacement to report the results.[20] The above-mentioned matching procedure results in 710 treated and 3,235 matched control consumers.[21] In Table 3, we report the results of our propensity score matching model: before matching, consumers with younger age, higher spending amount, more purchases, and more pageviews are more likely to adopt Tmall Genie; and females are also more likely to be adopters. Figure 2 shows that, after matching, both the treated and control users are similar in their adoption propensities. In Table 2(b), both t-test and Kolmogorov–Smirnov test results show that

[20] For robustness, we also conduct one-to-one matching with and without replacement and nearest three neighbors with and without replacement. Qualitatively similar results are found.
[21] 3235 < 710*5 because some adopters share overlapping matched non-adopters.

after matching, the two groups are indistinguishable in terms of their pre-adoption demographics and shopping behaviors.

Table 2. Comparison of Adopters and Non-Adopters Before and After Matching

	Adopter (N)	Adopter (Mean)	Non-adopter (N)	Non-adopter (Mean)	p-value of t-test	p-value of KS-test
(a) *Before matching*						
Age	710	30.77	39,175	32.06	0.000	0.000
Gender (Male = 1)	710	0.380	39,175	0.582	0.000	0.000
Total Spending	710	35.93	39,175	19.36	0.000	0.000
Total Pageviews	710	218.3	39,175	139.4	0.000	0.000
# Historical Purchases	710	569.9	39,175	290.2	0.000	0.000
(b) *After matching*						
Age	710	30.77	3,235	30.94	0.560	0.813
Gender (Male = 1)	710	0.380	3,235	0.387	0.740	0.999
Total Spending	710	35.93	3,235	34.34	0.599	0.073
Total Pageviews	710	218.3	3,235	220.8	0.832	0.124
# Historical Purchases	710	569.9	3,235	564.7	0.858	0.352

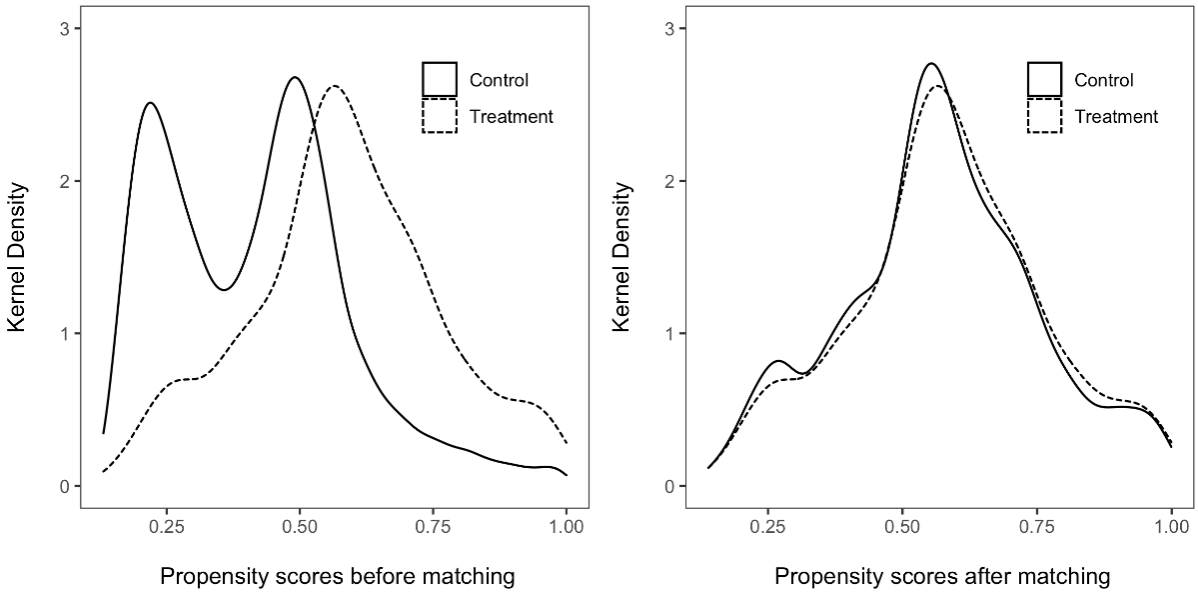
Notes: Calculated over the pre-adoption period. The unit of analysis is an individual user in (a) the unmatched, and (b) the matched sample. *p<0.05; **p<0.01; ***p<0.001.

Table 3. Propensity Score Model

Covariates	Coefficient
Age	-0.0274*** (0.0010)
Gender	-1.1211*** (0.0230)
Total Spending	0.0001*** (7.61e-6)
Total Pageviews	0.0004*** (2.66e-5)
# Historical Purchases	0.0007*** (1.4e-5)
Observations	39,885

Notes: Logit model with standard errors in parentheses. Estimates are calculated on an unmatched sample of 710 adopters and 39,175 non-adopters in the pre-adoption period; the unit of analysis is an individual user. The dependent variable is whether a user adopted Tmall Genie (adoption = 1), or not (adoption = 0). *p<0.05; **p<0.01; ***p<0.001.

Figure 2. Distribution of Propensity Scores Before and After Matching



4.4. Main Effect using Difference in Differences

We employ the difference in differences (DiD) model on matched individuals to estimate the treatment effect of Tmall Genie adoption on consumers' purchase and pageview behavior. The model for estimating the main effect on purchase (Model 1) and the main effect on pageview (Model 2) is specified as follows:

$$y_{it} = \alpha \cdot \text{Adoption}_i + \beta \cdot \text{Post}_t + \gamma \cdot (\text{Adoption}_i \times \text{Post}_t) + \delta_i + \varepsilon_{it}, \qquad (2)$$

where y_{it} is the total spending amount or the total pageviews of consumer *i* in week *t*, γ is the DiD estimator capturing the main effect. δ_i is a consumer-level fixed effect and ε_{it} is the error. The time fixed effect was not included because it is collinear with the Post_t term. We report robust standard errors clustered at the user level.

In summary, we combine the propensity score matching method (to select non-adopters who are like adopters) and the DiD approach. We assume that the unobserved ϵ_i in the propensity score model is independent of the unobserved ε_{it} in the DiD regression model.

4.5. Heterogeneity in Treatment Effect across User Demographics

The impact of voice shopping may depend on individual characteristics. From the company's perspective, knowing how Genie affects consumers with different demographics can help managers better target the users for Genie adoption. In this section, we consider how two potential user-level demographics moderate the treatment effect of Genie adoption on consumers' purchase and pageview behavior.

First, we consider the moderating effect of age. This is theoretically indeterminate. It could be that younger people are more open to new technology and thus change shopping behaviors more easily. On the other hand, younger consumers are more income constrained and would therefore make less change in shopping after the adoption of Tmall Genie. We empirically estimate the moderating effect of age on adoption in Model 3 (purchase) and Model 4 (pageviews)

$$\begin{aligned} y_{it} = \alpha \cdot \text{Adoption}_i + \beta \cdot \text{Post}_t + \lambda_{\text{age1}} \cdot (\text{Adoption}_i \times \text{Post}_t \times \text{I}[\text{Age}_i < 25]) \\ + \lambda_{\text{age2}} \cdot (\text{Adoption}_i \times \text{Post}_t \times \text{I}[25 \leq \text{Age}_i < 35]) + \lambda_{\text{age3}} \\ \cdot (\text{Adoption}_i \times \text{Post}_t \times \text{I}[35 \leq \text{Age}_i < 45]) + \lambda_{\text{age4}} \\ \cdot (\text{Adoption}_i \times \text{Post}_t \times \text{I}[\text{Age}_i \geq 45]) + \delta_i + \varepsilon_{it}, \end{aligned} \tag{3}$$

where y_{it} is the total spending amount or number of pageviews of consumer *i* in week *t*, Age_i is the age of consumer *i*, the indicator variable $\text{I}[\text{Age}_i < 25]$ would be 1 if the consumer is under 25 years old. λ is the DiD estimator capturing the moderating effect.

Similarly, the effect of adopting Tmall Genie on shopping behaviors may differ across genders. As it is also theoretically ambiguous, we estimate the moderating effect of gender with Model 5 (purchase) and Model 6 (pageviews) as specified below:

$$
\begin{aligned}
y_{it} = \alpha \cdot \text{Adoption}_i + \beta \cdot \text{Post}_t + \lambda_{\text{male}} \\
\cdot (\text{Adoption}_i \times \text{Post}_t \times \text{I}[\text{Gender}_i = \text{'male'}]) + \lambda_{\text{female}} \\
\cdot (\text{Adoption}_i \times \text{Post}_t \times \text{I}[\text{Gender}_i = \text{'female'}]) + \delta_i + \varepsilon_{it}
\end{aligned} \tag{4}
$$

where y_{it} is the total spending amount or number of pageviews of consumer *i* in week *t*, Gender_i is the gender of consumer *i*. The indicator variable $\text{I}[\text{Gender}_i = \text{'female'}]$ would be 1 if the consumer is a female. λ is the DiD estimator capturing the moderating effect.

4.6. Short-term versus Long-term Effect

To investigate how long the changes in purchase and browsing last, we also distinguish between treatment effects in the short term (within the first week of adoption, λ_{ST}), medium term (between 2 and 7 weeks after adoption, λ_{MT}), and long term (8 weeks and after, λ_{LT}). We estimate the following Model 7 (purchase) and Model 8 (pageviews).

$$
\begin{aligned}
y_{it} = \alpha \cdot \text{Adoption}_i + \beta \cdot \text{Post}_t + \lambda_{\text{ST}} \\
\cdot (\text{Adoption}_i \times \text{Post}_t \times \text{I}[0 \leq \text{week\_since\_adoption}_t \leq 1]) + \lambda_{\text{MT}} \\
\cdot (\text{Adoption}_i \times \text{Post}_t \times \text{I}[2 \leq \text{week\_since\_adoption}_t \leq 7]) + \lambda_{\text{LT}} \\
\cdot (\text{Adoption}_i \times \text{Post}_t \times \text{I}[\text{week\_since\_adoption}_t \geq 8]) + \delta_i + \varepsilon_{it}
\end{aligned} \tag{5}
$$

where y_{it} is the total spending amount or number of pageviews of consumer *i* in week *t*, and $I[\text{week\_since\_adoption}_t]$ is the indicator variable.

4.7. Heterogeneity in Treatment Effect across Channels

With the introduction of voice as a new channel of shopping, a natural question is how it affects other channels (e.g., PC and mobile). Would the voice channel cannibalize consumers' purchase from other channels, or would it help expand purchase/browsing for all channels? We examine the issue in this subsection.

The model is the same as that used to estimate the main effect except that we use different subsets of purchase data for each channel: Model 9 (mobile channel), Model 10 (PC channel), and Model 11 (voice channel). As our pageview data does not record the channel information, the analysis can only be done for consumers' purchase behavior.

4.8. Heterogeneity in Treatment Effect across Categories

In this subsection, to further understand how Tmall Genie adoption affects consumers shopping behavior for different types of products, we investigate the treatment effect on purchase across product categories.

Prior research (Munz and Morwitz 2020) has found that information presented by voice can be more difficult to process than the same information presented in writing. And auditory consumers who shop by voice are less able to differentiate between choice options. Therefore, consumers prefer to use voice shopping for products that do not require active searching or comparison. Based on this reasoning, we develop and test two hypotheses using data on the user-week-category level in this subsection.

Hypothesis 1. The impact of voice AI adoption on purchase is stronger for categories with lower substitutability.

$$y_{itj} = \alpha \cdot \text{Adoption}_i + \beta \cdot \text{Post}_t + \gamma \cdot (\text{Adoption}_i \times \text{Post}_t) + \lambda_{\text{LS}} \cdot \left(\text{Adoption}_i \times \text{Post}_t \times \text{LowSubstitutability}_j\right) + \delta_i + \sigma_j + \varepsilon_{itj}, \tag{6}$$

where y_{itj} is the spending amount of consumer i in week t on category j, $\text{LowSubstitutability}_j$ is a dummy variable which would be equal to 1 if j belongs to categories with lower substitutability and 0 otherwise. γ is the DiD estimator capturing the main effect. λ_{LS} is the DiD estimator capturing the interaction effect. The way we evaluate the substitutability of a product category is

by going through existing findings in the literature. For example, prior literature has found that the substitutability of video games (Lee 2003, Nair 2007), books (Chevalier and Goolsbee 2003, McMillan 2007), movies (Leslie, 2004; Orbach, Einav, 2006), music (Shiller and Waldfogel, 2009), sports ticket (Luini and Sabbatini 2010), is low such that each product can be treated as an independent market. Intuitively, a consumer who wants to watch a baseball game will not substitute a baseball ticket with a basketball ticket. We test this hypothesis by estimating two models using different subsets of the purchase data: Model 12 (using all-channel purchase data) and Model 13 (using only voice-channel purchase data).

In addition to substitutability, purchase frequency can be another moderator. The intuition is that if consumers frequently purchase in a specific category, then they are very familiar with the products in this category. So no active search or comparison is necessary prior to each new purchase any more. Therefore, consumers are more likely to use Genie to purchase products in such a category. We proceed to examine the following hypothesis concerning categories of different purchase frequency.

Hypothesis 2. The impact of voice AI adoption on purchase is stronger for categories with high purchase frequency.

To test this hypothesis, we define frequency as the number of purchases made by consumers at the week level. We then add the interaction between purchase frequency (as a continuous variable) of each category and $\text{Adoption}_i \times \text{Post}_t$:

$$y_{itj} = \alpha \cdot \text{Adoption}_i + \beta \cdot \text{Post}_t + \gamma \cdot (\text{Adoption}_i \times \text{Post}_t) + \lambda_{\text{F}} \cdot \left(\text{Adoption}_i \times \text{Post}_t \times \text{Frequency}_j\right) + \delta_i + \sigma_j + \varepsilon_{itj}, \tag{8}$$

where y_{itj} is the spending amount of consumer i in week t on category j, Frequency_j is the purchase frequency of category j. γ is the DiD estimator capturing the main effect. λ_{F} is the DiD estimator capturing the interaction effect. Following the similar vein, we estimate Model 14 (using all-channel purchase data) and Model 15 (using only voice-channel purchase data) to show the hypothesis testing results.

5. Results

5.1. Main Effects

We first report the main effect of Genie adoption on consumers' purchase and pageview behavior, corresponding to Model 1 and Model 2 as specified in Section 4.

Table 4. Main Effects

Variable	(1) Model 1 (Purchase)	(2) Model 2 (Pageview)
Adoption × Post	8.413*** (2.461)	3.766 (3.517)
Post	-3.860*** (0.999)	-7.365*** (1.492)
Adoption	-4.885 (39.66)	79.32 (59.27)
Constant	35.51*** (0.639)	206.1*** (10.71)
Individual FE	Yes	Yes
Observations	122,295	122,295
Adjusted R^2	0.095	0.618

Note: Robust standard errors in parentheses.
*p<0.05; **p<0.01; ***p<0.001

In Table 4, Column (1), we can see that on average Tmall Genie adoption leads consumers to spend RMB 8.413 more in a week, which is a growth of 23% (= 8.413 / 35.93). Assuming that the Genie adopters in our randomly chosen sample are representative of the consumers adopting the

voice AI on Alibaba, our results suggest that the adoption of Genie can lead to around RMB 4.4 billion (= 8.413 × 52 weeks × 10 million adopters) (US$ 630 million) increase in annual sales every year. On the contrary, the effect on pageview is positive yet insignificant over the entire post-treatment period, as indicated in Column (2) of Table 4.

To sum up, Genie adoption leads to significant growth in spending for online shopping. This finding implies that having Genie can give consumers a higher tendency to make an order, possibly due to the higher level of convenience and more flexibility brought by the voice shopping assistant. Note that since we have data from only one e-commerce platform, the increased spending and browsing behaviors could be explained by consumer switching from alternative platforms to the focal platform or primary consumption expansion. Unfortunately, we cannot tease out these two possibilities due to data limitations.

5.2. Heterogeneous Treatment Effects across User Demographics

There are sizeable effects of adopting Genie. We now explore how these effects differ across adopters. In Table 5, we show the heterogeneous treatment effects for different user groups. We focus on two particular moderators: age and gender.

For age, we split users into four groups: younger than 25, between 25 and 35, between 35 and 45, and older than 45. The results in Column (1) of Table 5 indicate that the Genie adoption effect on purchase spending is the strongest (RMB 15.98 more, $p < 0.001$) for users younger than 25 years old, followed by the age group [35, 45) (RMB 6.979 more, $p < 0.05$) and the age group [25, 35) (RMB 13.03 more, $p < 0.05$). The effect on those older than 45 is negative (RMB 21.22 more, $p < 0.05$); a possible explanation is that this group of consumers find it less straightforward to use the voice shopping assistant, which limits the number of products they can consider and purchase. The results in Column (3) show that the effect of Genie adoption on pageviews is only significant for

consumers younger than 25 years old, with the effective size being (14.31 more pageviews, $p < 0.001$) per consumer per week.

Table 5. Heterogeneous Treatment Effect across User Demographics

Variable	(1) Model 3 (DV = Purchase)	(2) Model 4 (DV = Purchase)	(3) Model 5 (DV = Pageview)	(4) Model 6 (DV = Pageview)
Adoption × Post × Age < 25	15.98*** (4.297)		14.32* (6.421)	
Adoption × Post × Age in [25, 35)	6.979* (2.990)		-1.539 (4.469)	
Adoption × Post × Age in [35, 45)	13.03* (5.389)		9.638 (8.053)	
Adoption × Post × Age > 45	-21.22* (8.620)		23.21 (12.88)	
Adoption × Post × Male		10.28** (3.568)		-12.63* (5.331)
Adoption × Post × Female		6.989* (2.915)		16.04*** (4.356)
Adoption	-4.052 (39.67)	-5.967 (39.69)	82.40 (59.29)	88.84 (59.31)
Post	-3.959*** (0.994)	-3.824** (0.998)	-7.739*** (1.486)	-7.556*** (1.491)
Constant	35.36*** (7.173)	35.70*** (7.177)	205.5*** (10.72)	204.4*** (10.72)
Individual FE	Yes	Yes	Yes	Yes
Observations	122,295	122,295	122,295	122,295
Adjusted R^2	0.096	0.095	0.605	0.605

Note: Robust standard errors in parentheses.
*p<0.05; **p<0.01; ***p<0.001

We also examine the treatment effect of Genie adoption on female and male consumers separately. We find that the increase in purchase of male consumers is larger than that of female consumers (RMB 10.28 > RMB 6.989). We also find that the treatment effect on browsing for female is

positive and significant (16.04 more pageviews. $p < 0.001$), whereas for male it is negative and significant (12.63 less pageviews. $p < 0.05$).

An implication of the above findings is that different user groups may incur different behavior changes after adopting Tmall Genie. But overall the effects on purchase are positive and significant, which is consistent with our main findings. In contrast, the increase in pageviews is correlated with age and gender, that is, younger consumers and female tend to browse more as the result of adopting voice AI. Our findings also indicate that browsing more does not necessarily mean purchasing more. Managers can refer to our findings of the heterogeneous treatment effect across demographics to guide product targeting strategies.

5.3. Short-term versus Long-term Impact

How long do the changes in purchase and browsing last? In Table 6, we report results of treatment effects in the short term (within the first week after adoption), medium term (between 2 and 7 weeks after adoption), and long term (8 weeks and after). Column (1) corresponds to the model with purchase spending as the dependent variable, whereas Column (2) speaks to the results on pageviews.

We find that the treatment effect on purchase is the highest (RMB 13.94 more) for the first week after adopting Tmall Genie. The effects shrink to smaller magnitudes in the mid-term (RMB 8.566 more) and long-term (RMB 7.216 more), but still remain positive and significant. On the contrary, the effect on browsing is only significant in the short-term (35.76 more pageviews) and vanish in the long run. It is likely that, shortly after adopting the voice assistant, consumers have some curiosity or urge to go through some trials to get familiar with the newly adopted AI gadget. As a result, the increase in purchase is the highest compared to later stages.

It is noteworthy that the significant positive effect on purchase never vanished in all terms. From midterm to long-term, the decrease in magnitude is small (from RMB 8.566 to RMB 7.216) for purchase change. To managers, this is reassuring that the voice assistant device not only creates more purchase interest in the short run but also retains such interest in the long run, thus it could be that the voice assistant is not a gimmick but a product worth future development and investment.

Table 6. Short-term versus Long-term Effect

Variable	(1) Model 7 (DV = Purchase)	(2) Model 8 (DV = Pageview)
Adoption × Post × Short-term	13.94** (4.559)	35.76*** (6.812)
Adoption × Post × Mid-term	8.566** (3.058)	-9.019 (4.815)
Adoption × Post × Long-term	7.216** (2.658)	5.038 (3.972)
Adoption	-4.885 (39.65)	79.33 (59.26)
Post	-3.860*** (0.999)	-7.365*** (1.492)
Constant	35.51*** (7.171)	206.1*** (10.71)
Individual FE	Yes	Yes
Observations	122,295	122,295
Adjusted R^2	0.095	0.605

Note: Robust standard errors in parentheses.
*p<0.05; **p<0.01; ***p<0.001

5.4. Treatment Effect at Channel Level

The findings presented so far indicate a positive effect on consumer purchase and browsing. We proceed to examine how the voice channel would affect demand made through other channels. Would the demand increase for the whole online shopping market or just switching from one channel to the other? Given we have the channel information for each completed order (but not for

pageview), we estimate the treatment effect on purchase separately for each channel. The results are displayed in Table 7.

First of all, we can see the signs of treatment effects for the three channels are all positive, meaning there is no cannibalization from the voice channel on the other two traditional channels. Instead, voice shopping has a positive spillover effect on other shopping channels. Mobile channel benefits the most (effect size = RMB 4.420, $p < 0.05$), whereas the demand through PC channel is not significantly affected. This is reasonable as one limitation of the mobile channel is found to be the small screen size that displays less product options compared to PC, which makes search harder (e.g., Chae and Kim 2004; Ghose et al. 2013). With the smart voice assistant, consumers can narrow down the consideration set first, and then they may feel more comfortable making the payment or purchase decision on their mobile phones.

Table 7. Treatment Effect at Channel Level

Variable	Model 9 (Mobile Channel)	Model 10 (PC Channel)	Model 11 (Voice Channel)
Adoption × Post	4.420*	1.224	2.769***
	(2.087)	(1.036)	(0.084)
Adoption	-2.567	-0.711	-1.608
	(35.17)	(17.45)	(1.423)
Post	-1.971*	-1.891***	0.001
	(0.886)	(0.439)	(0.036)
Constant	28.47***	6.669*	0.367
	(6.359)	(3.156)	(0.257)
Individual FE	Yes	Yes	Yes
Observations	122,295	122,295	122,295
Adjusted R^2	0.092	0.067	0.070

Note: Robust standard errors in parentheses.
*p<0.05; **p<0.01; ***p<0.001

5.5. Treatment Effect across Categories

In Section 4.8, we propose two hypotheses about how voice AI adoption affects consumers' shopping behavior for different types of products, one concerning products' substitutability (Model 12 and 13) and the other about the purchase frequency (Model 14 and 15). The findings are shown in Table 8.

Table 8. Empirical Evidence to Support Hypotheses

Variable	*Hypothesis 1*		*Hypothesis 2*	
	(1) Model 12 (all channel)	(2) Model 13 (voice channel)	(3) Model 14 (all channel)	(4) Model 15 (voice channel)
Adoption	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)	0.000 (0.000)
Post	-0.027** (0.007)	0.000 (0.000)	-0.027** (0.007)	0.000 (0.000)
Adoption × Post	-0.005 (0.016)	-0.003*** (0.001)	-0.015 (0.017)	-0.003*** (0.001)
Adoption × Post × Low Substitutability	2.285*** (0.063)	0.791*** (0.002)		
Adoption × Post × Purchase Frequency			20.89*** (0.689)	395.9*** (0.561)
Constant	0.242*** (0.005)	0.001** (0.001)	0.242*** (0.005)	0.001 (0.001)
Individual FE	Yes	Yes	Yes	Yes
Category FE	Yes	Yes	Yes	Yes
Observations	17,488,185	17,488,185	17,488,185	17,488,185
Adjusted R^2	0.011	0.011	0.011	0.032

Note: Robust standard errors in parentheses.
*p<0.05; **p<0.01; ***p<0.001

Comparing the all-channel and voice-channel results, we can see the signs and significance levels are all the same. As our hypothesis is more related to how the voice shopping experience differs across product categories, we focus on the interpretation of the voice-channel results (i.e., Column (2) and (4)).

The estimate of the coefficient for the interaction term Adoption × Post × Low Substitutability in Column (2) shows support for *Hypothesis 1*; that is, the impact of voice AI adoption on purchase is stronger for categories with lower substitutability. The rationale is that consumers who shop by voice are less able to differentiate between choice options because information presented by voice can be more difficult to process than the written information. Therefore, consumers are more likely to use the voice channel for products that do not require active searching or comparison. Since low substitutability reduces the need to compare between products before making a purchase, we see a stronger treatment effect for the corresponding categories.

Similarly, the results in Column (3) and (4) show support for *Hypothesis 2*: the impact of voice AI adoption on purchase is stronger for categories with high purchase frequency. Because consumers have frequently purchased from these products, they are very familiar with the products characteristics and the usage experience. Therefore, there is no need to actively search or compare between products when they repetitively make purchases. As a result, there is a stronger effect for categories with higher purchase frequency.

5.6. Robustness Checks

5.6.1. Placebo Treatments

As the DiD approach requires the assumption of parallel pre-treatment trends, we test whether it holds by carrying out a "placebo" treatment test. In particular, we define a set of placebo "treatments" at some midpoints of the pre-adoption period and conduct sample matching using such placebo settings. Next, we estimate the same DiD model to examine the main effect. Theoretically, this treatment effect should be insignificant as those placebo "treatments" are not real. The empirical evidence fails to reject the null hypothesis of no treatment effect for the placebo

"treatments" (see results in the Appendix Table A2). This suggests that pretreatment trends are statistically equivalent across both user groups.

5.6.2. Controlling for Technology Enthusiasm

Although we have matched individuals on their historical shopping and browsing behaviors, someone might be worried that Genie adopters may have stronger enthusiasm for new technology (such as the voice assistant) than non-adopters, and this technology enthusiasm cannot be reflected and controlled for by merely looking at historical shopping and browsing records across all categories. As a check against this, we add the historical purchase and pageviews on the "smart devices" category – which controls for a user's preference for new consumer technologies – as additional covariates in the propensity score matching step and replicate the entire analysis. The results using this newly matched sample remain qualitatively similar to the main results.

5.6.3. Removal of Outliers

Our third robustness check involves the sensitivity analysis with the outliers in the purchase and browsing data removed, as someone might have concerns that the effect may be due to the outliers in data. For example, the maximum observation for the total spending amount on the user-week level as shown in Table 1 is RMB 246,251 (US$ 36,900), which should be deemed uncommon. To perform this check, we remove those purchase records whose spending amounts are greater than the 99.9 percentile and replicate the whole analysis. Qualitatively similar findings are derived, which proves that our estimation of the effects is unlikely to suffer from this data issue.

5.6.4. Evidence using an Additional Dataset

As described in subsection 3.1, a potential concern with our empirical setting is that consumers may choose to adopt Genie due to some shocks that are not random. Even though we have confirmed with Alibaba that those adopters adopted Genie without any promotion activities and

the placebo test serves as additional evidence, someone might still have other concerns about the robustness of our findings. In the online appendix, we provide extra empirical evidence using another dataset collected in a different time period from Alibaba. The findings are qualitatively consistent with the present ones.

6. Implications and Conclusions

Voice shopping is gaining growing significance in our daily lives and the global economy. This paper takes the first step towards understanding the causal impact of voice shopping on consumers' shopping behavior. We first summarize the main findings yielded by our analysis: *First*, we find that consumers on average spend 23% more as the result of adopting the voice AI, which translates to an approximately RMB 4.4 billion (US$ 630 million) increase in annual sales revenue every year. *Second*, we find heterogeneous treatment effects for different groups of consumers: the effect on purchase is stronger for younger consumers and male, and the effect on browsing is more salient for younger consumers and female. *Third*, we find that although the positive effect on purchases shrinks over time, it still remains significant and large after eight weeks. *Fourth*, the voice channel does not cannibalize other purchase channels (such as the PC channel and the mobile channel); rather, it has a positive spillover effect on the mobile channel. *Lastly*, the impact of voice AI adoption on purchase is stronger for products that do not require active search or comparison, such as product categories with low substitutability or high purchase frequency. To summarize, we demonstrate that the adoption of voice AI has enhanced the growth of the digital sales market in the world's largest e-commerce platform.

This study provides several managerial implications. First, our research speaks to the e-commerce firms on how they can capitalize on the emerging voice AI technology. Voice activated devices have been serving multiple needs of customers, such as making calls, setting alarms, playing music,

and controlling other devices. What makes e-commerce's voice-activated shopping special, especially in a world where IT giants like Apple and Google are major players, is its rich connection with a customer's purchase history and the ability to allow shopping in the large marketplace by voice. Given neither consumer shopping records nor the voice AI is transferable across firms, e-commerce firms may want to develop their own voice shopping assistant, for stimulating higher volume of sales and search. Moreover, according to our findings on heterogeneous treatment effect of the voice AI adoption, managers may want to promote the voice AI device to consumers who tend to spend, purchase, or search more after using the smart device.

Second, our results show that the voice shopping channel has positive spillover effect on the PC and mobile channels in terms of the spending amount. One possible cause of this complementary relationship could be that consumers use the voice AI and other channels together to make a purchase. Managers may want to keep this positive spillover effect in mind when trying to engage consumers through platform design. For example, e-commerce sites can keep records of where consumers ended in the voice search channel, and then show them to consumers when they open up the site on mobile or PCs.

Third, our findings indicate that the adoption of the voice AI can induce larger increases in sales and search for categories with high purchase frequency or low substitutability. There are two immediate implications to the e-commerce firm. First, marketers may want to promote the voice AI together with product categories that are frequently purchased or of low substitutability. Second, the firm can try to target the voice AI to consumers who buy a lot from product categories with high purchase frequency or low substitutability. By doing as suggested, managers can save on marketing spending and achieve a higher return of investment.

There are some remaining questions beyond the scope of this paper but can be venues for future research. First, as our data does not cover consumers' purchase and search behavior on other e-commerce platforms, we cannot measure the overall effect of voice AI adoption on the entire e-commerce market. Our findings indicate market expansion benefits for the examined platform only, but it is unclear whether the competitors will face a loss or not. Second, our findings are derived from data within a time window in 2019. As consumers in the future get more and more familiar with artificial intelligence and related products, the magnitude of the impact of voice AI adoption on consumers' shopping behavior may change. Our findings on the short-term and long-term effects partially address this issue but caution is warranted in understanding the effect in an even longer term. Lastly, in this study we keep ignorant about how a voice AI can better serve consumers. Future research can explore this research direction if richer data containing the conversation between the voice AI and consumers are obtained.

References

Ba W, Mendelson H, Zhu M (2020) Sales Policies for a Virtual Assistant. ArXiv preprint arXiv: 2009.03719.

Bronnenberg B, Dubé JP, Mela C (2010) Do digital video recorders influence sales? *Journal of Marketing Research* 47(6):998–1010.

Chae M, Kim J (2004) Do size and structure matter to mobile users? An empirical study of the effects of screen size, information structure, and task complexity on user activities with standard web phones. *Behaviour & Information Technology*. 23(3):165–181.

Chevalier J, Goolsbee A (2003) Measuring prices and price competition online: Amazon. com and BarnesandNoble. com. *Quantitative marketing and Economics*, 1(2), 203-222.

Datta H, Knox G, Bronnenberg BJ (2018) Changing their tune: How consumers' adoption of online streaming affects music consumption and discovery. *Marketing Science*, 37(1), 5–21.

Dawar N, Bendle N (2018) Marketing in the age of Alexa. *Harvard Business Review*, 96(3), 80-86.

Ghose A (2018) TAP: Unlocking the mobile economy. MIT Press.

Ghose A, Goldfarb A, Han SP (2013) How is the mobile Internet different? Search costs and local activities. *Information Systems Research*, 24(3), 613-631.

Jones VK (2018) Voice-activated change: Marketing in the age of artificial intelligence and virtual assistants. *Journal of Brand Strategy*, 7(3), 233-245.

Kumar V, Dixit A, Javalgi RRG, Dass M (2016) Research framework, strategies, and applications of intelligent agent technologies (IATs) in marketing. *Journal of the Academy of Marketing Science*, 44(1), 24-45.

Lee RS (2013) Vertical integration and exclusivity in platform and two-sided markets. *American Economic Review*, 103(7), 2960-3000.

Leslie P (2004) Price Discrimination in Broadway Theatres, *RAND Journal of Economics*, 35, 3, 520-41

Luini L, Sabbatini P (2010) Demand cross elasticity without substitutability: Evidence from an experiment. Available at SSRN 1617337.

Luo X, Qin MS, Fang Z, Qu Z (2020) Artificial Intelligence Coaches for Sales Agents: Caveats and Solutions. *Journal of Marketing*, forthcoming.

Luo X, Tong S, Fang Z, Qu Z (2019) Frontiers: Machines vs. humans: The impact of artificial intelligence chatbot disclosure on customer purchases. *Marketing Science*, 38(6), 937-947.

Mari A, Mandelli A, Algesheimer R (2020) The evolution of marketing in the context of voice commerce: A managerial perspective. In International Conference on Human-Computer Interaction (pp. 405-425). Springer, Cham.

McMillan R (2007) Different Flavor, Same Price: The Puzzle of Uniform Pricing for Differentiated Products, mimeo, Federal Trade Commission

Munz K, Morwitz V (2020) Not-so Easy Listening: Roots and Repercussions of Auditory Choice Difficulty in Voice Commerce. Available at SSRN 3462714.

Nair H (2007) Intertemporal price discrimination with forward-looking consumers: Application to the US market for console video-games. *Quantitative Marketing and Economics*, 5(3), 239-292.

Orbach BY, L Einav (2007) Uniform Pricing for Differentiated Goods. The Case of Movie Theatre Industry, *International Review of Law and Economics*, 27, 129-153

Shiller BJ, Waldfogel (2009) Music for a Song: An Empirical Look at Uniform Song Pricing and its Alternatives, NBER Working Paper: 1539

Smith KT (2020) Marketing via smart speakers: what should Alexa say? *Journal of Strategic Marketing*, 28(4), 350-365.

Tan TF, Netessine S (2020) At your service on the table: Impact of tabletop technology on restaurant performance. *Management Science*, 66(10), 4496-4515.

Wang RJH, Malthouse EC, Krishnamurthi L (2015) On the go: how mobile shopping affects customer purchase behavior. *Journal of Retailing*, 91(2), 217–234.

Xu J, Forman C, Kim JB, Van Ittersum K (2014) News media channels: Complements or substitutes? Evidence from mobile phone usage. *Journal of Marketing*, 78(4), 97–112.

Xu K, Chan J, Ghose A, Han SP (2017) Battle of the channels: The impact of tablets on digital commerce. *Management Science*, 63(5), 1469-1492.

Xu Y, Ghose A, Xiao B (2019) Mobile Payment Adoption: An Empirical Investigation on Alipay. Available at SSRN 3270523.

Appendix

The Contemporary Use of Major Voice-Devices

The focal voice assistant device being investigated in this paper is Alibaba Tmall Genie, which is the No.1 voice AI device in China. Feature-wise[22], Alibaba is offering much of the same functionality as the Amazon Echo, including smart home control, weather, news, music, and a wide range of other skills. These can all be activated by saying "Tmall Genie" in Mandarin. As the name suggests, the Tmall Genie can order items from Tmall, Alibaba's shopping site, through voice commands; it has voiceprint recognition to ensure that only authorized users can place orders. It has been marketed as a shopping device. The table[23] below compares the features of major voice devices in the market.

Major Devices	Tmall Genie	Amazon Echo	Google Home	Apple Homepod
Shopping	✓	✓	✓	×
Track packages	✓	✓	×	×
Web search	✓	✓	✓	✓
Smart home control[24]	✓	✓	✓	✓
News/Music	✓	✓	✓	✓
Phone call	✓	✓	✓	✓
Skills[25]	✓	✓	✓	✓
Read audiobook	✓	✓	✓	×
Intercom[26]	×	×	✓	×

[22] Cited from: https://www.theverge.com/circuitbreaker/2017/7/5/15921722/alibaba-amazon-echo-competitor-tmall-genie-beta-voice-speaker-shop.
[23] The unique features of each device can be found here: https://thewiredshopper.com/google-home-vs-amazon-echo-vs-apple-homepod/#tab-con-11. We also leveraged other online resources to create the table.
[24] Some smart home devices are only compatible with some of the smart speakers. For example, thermostats and security cameras from Nest, as well as Samsung SmartThings, are only compatible with Amazon Echo and Google Home.
[25] Skills are called "Actions" for Google Home, and "Abilities" for Apple Homepod.
[26] Cited from: https://www.forbes.com/sites/forbes-personal-shopper/2019/05/14/apple-homepod-vs-amazon-echo-vs-google-home-which-smart-speaker-is-right-for-you/#15dd19df10ef

Institutional Details of the Voice Shopping Process

We provide more institutional details of how to use Tmall Genie to make purchases in this section. Here are a few stylized facts. First, the voice AI recommends products to consumers and Alibaba currently uses the same search engine ranking algorithm to sort products based on the specific queries a consumer uses, no matter if the query comes from the web interface or from voice AI. For example, below is a representative conversation.

Customer	"Tmall Genie, I wanted to buy potato chips."
Genie	"Master, I recommend Dali Park Kebby potato chips four packs, each pack 32 grams, free shipping, actual price 9.98 yuan. It will be delivered to (address), may I place the order for you?"
Customer	"Yes, please place the order."
Genie	"Sure! In order to proceed, let us do voice authentication first. Please keep quiet around, and after the 'beep', say 'Tmall Genie, 9951.'" (Here 9951 is the authentication code randomly generated by the system.)
Customer	"9951".
Genie	"Alipay discount is applied. If you want to know the delivery status, you can let me know by saying 'Tmall Genie, tracking information.'"

Another example when the consumer specifies the brand is as follows.

Customer	"Tmall Genie, I wanted to have some Lay's potato chips."
Genie	"Master, I recommend Lay's three-piece canned snack food package, shipping cost 20 yuan, actual price 37.9 yuan. It will be delivered to (address), may I place the order for you?"
Customer	"Not really this time."
Genie	"Ok, looking forward to your next use."

In both examples, there is only one recommended product in each answer the voice AI provides. And the recommended product is the same as the first result listed on the Tmall website when the consumer conducts text-based search. For instance, in the first example, when the consumer types "potation chips" in the search bar on the Tmall website, the first returned result is the "Dali Park

Kebby potato chips four packs, each pack 32 grams, free shipping, actual price 9.98 yuan." And in the second example, when the consumer enters "Lay's potato chips" in the search box, the first returned result is the "Lay's three-piece canned snack food package, shipping cost 20 yuan, actual price 37.9 yuan." And the second result is the "Lay's black pepper cheese non-fried potato chips shipping cost 20 yuan, actual price 37 yuan."

Second, if the consumer does not fully specify the product, e.g., "order a refill card" instead of also specifying the amount, the voice AI will recommend a product with a specific payment amount. Below is the example when a consumer orders a refill card without specifying the amount.

Customer	"Tmall Genie, I'd like to order a mobile refill card."
Genie	"Master, I would recommend China Mobile's refill card. The total price is 100 RMB. It will be delivered to (address). May I place the order for you?"
Customer	"Yes, please place the order."
Genie	"Sure! In order to proceed, let us do voice authentication first. Please keep quiet around, and after the 'beep', say 'Tmall Genie, 2065.'" (Here 2065 is the authentication code randomly generated by the system.)
Customer	"2065".
Genie	"Alipay discount is applied. If you want to know the delivery status, you can let me know by saying 'Tmall Genie, tracking information.'"

Similar to the two examples provided before, the recommended product is the No.1 returned result when the search query "refill card" is entered on the Tmall website.

Figure A1 Adoption Rate across Age and Gender Observed in the Treatment Group

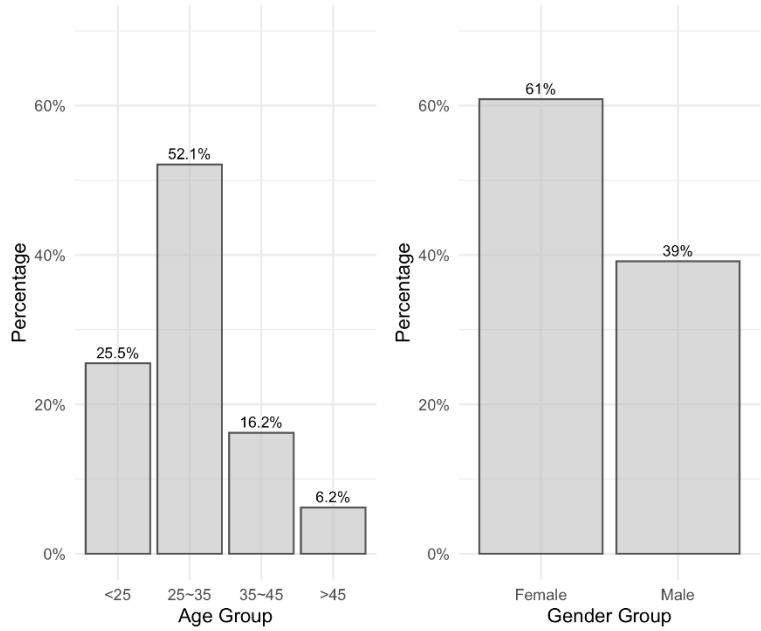


Table A1 Bayesian Posterior Adoption Rate

	Observation	Prior	Posterior
Adoption		2.15%	
Female	61%	58.64%	2.24%
Male	39%	41.36%	2.23%
Age < 25	25.5%	43.75%	1.25%
Age in [25, 35)	52.1%	44.39%	2.52%
Age in [35, 45)	16.2%	7.66%	4.55%
Age > 45	6.2%	4.2%	3.17%

Note: We employ Bayesian rule to derive the posterior estimation of the adoption rate across gender and across age groups. For example, Pr (Adoption | Male) = Pr (Male | Adoption) * Pr (Adoption) / Pr (Male). The observation data, namely Pr (Male | Adoption) in the first column, is deduced from our data, whereas the prior, namely Pr (Adoption) in the first row, along with Pr (Male) are from consulting firms. For example, some sample statistics can be found at: https://n.znds.com/article/37815.html and https://www.iimedia.cn/c1061/69035.html. The posterior probability can inform us about the adoption rate in a more intuitive way than the observed likelihood. For example, in order to answer the question of what is the adoption rate of male vs. female, one needs to report Pr (Adoption | Male) = 2.23% instead of Pr (Male | Adoption) = 39%. The last column of this table can be used to answer the question of this kind.

Table A2 Main Effects using Placebo Treatments

Variable	Placebo Tr. at Week 20th		Placebo Tr. at Week 21st	
	Model 1 (Purchase)	Model 2 (Pageview)	Model 1 (Purchase)	Model 2 (Pageview)
Adoption × Post	2.860	5.651	4.006	7.998
	(2.845)	(6.974)	(2.853)	(6.410)
Post	2.101	-4.848	2.109	-4.897
	(1.209)	(3.581)	(1.630)	(3.646)
Adoption	1.691	12.65	1.725	12.33
	(2.168)	(56.91)	(2.458)	(60.30)
Constant	27.04***	240.9***	25.58***	236.7***
	(7.665)	(13.48)	(6.814)	(13.41)
Individual FE	Yes	Yes	Yes	Yes
Observations	122,295	122.295	122,295	122.295
Adjusted R^2	0.090	0.610	0.091	0.610

Note: *p<0.05; **p<0.01; ***p<0.001

Table A3 List of All Categories

Category
Agriculture Equipment
Anime
Art & Collectibles
Auction
Audio Visual Electronics
Baby Diapers
Baby Dietary Supplement
Baby Products
Bags
Bath & Beauty
Bedding
Bicycles & Accessories
Books
Building Material
Cainiao Shop
Cameras
Car Care & Electronics
Car Maintenance Services
Cars
Cash Cards
Charity
Child Care Supplies
Children's Shoes
Clean Energy Transportation
Cleaning Supplies
Coffee & Beverages
Commercial Appliances
Computers & accessories
Contact lenses
Cosmetics
Coupons
Craft Supplies
Custom Design
Data Storage
Desktop Computers
Dietary Supplements
Digital Reading
DIY PC
Donation
Door-to-door Services
Dry Goods & Seasoning
Education & Training
Ele.me Delivery Services
Electric Vehicle
Electrical Engineering
Electronic Components
Electronics
Electronics Accessories
Entertainment Gift Card
Farm & Agricultural Supplies
Fashion Accessories
Flower Services
Food Ware
Fresh Hema
Furniture
Game Time Cards
Gift Cards (Offline)
Government Assets
Graphic Design Services
Hair Care Products
Hardware & Tools
Health & Fitness
Holiday Shop
Home & Living
Home Appliances
Home Carpet & Flooring
Home Décor Fabric
Home Decoration
Hotel
House Painting & Decor
In-Game Item Marketplace
Insurance
Jewelry
Jewelry & Accessories
Kids' Clothing
Kitchen Appliances
Kitchen Supplies
Laptops
Latex Products
Lights
Local Services
Loungewear
Loyalty Card
Meat & Vegetables
Medical Services
Medication Delivery Services
Men's Clothing
Men's Shoes
Mobile Phone Plans
Mobile Phones
Mobile Refill Card
Motorcycles
Movies & Sports Ticket
Music & Movies
Music Instruments
Network Devices
O2O Store
Office & School Supplies
Office Equipment
Office Furniture
OTC Medicines
Others
Packaging
Personal care
Personalized Interior Design
Pet Supplies
Pocket Knives & Glasses
Pregnancy & Maternity Products
Prescription Drug
Restaurant Gift Cards
Retail Logistics Services
Sex Toys
Shopping Gift Cards
Smart Home
Snacks
Software
Spa
Spa Supplies
Sport shoes
Sporting Goods
Sports & Outdoors
Sports Accessories
Sports Clothing
Storage & Organizer
Store Coupons
Tablets
Taobao Gift Card
Tea
Tencent QQ
Tmall O2O retailer
Tours & Activities
Toys
Transportation
Travel Services
Used Electronics
Used Good
Video Game
Watches
Wellness
Wine & Liquor
Women's Clothing
Women's Shoes