

The Choice Overload Effect in Online Recommender Systems

Xiaoyang Long^{1*}, Jiankun Sun², Hengchen Dai³, Dennis J. Zhang⁴
Jianfeng Zhang⁵, Yujie Chen⁵, Haoyuan Hu⁵, Binqiang Zhao⁵

1. Wisconsin School of Business, University of Wisconsin-Madison, Madison, WI, USA

2. Imperial College Business School, Imperial College London, London, UK

3. Anderson School of Management, University of California, Los Angeles, Los Angeles, CA, USA

4. Olin Business School, Washington University in St. Louis, St. Louis, MO, USA

5. Alibaba Group, Hangzhou, China

Problem Definition: Online retailing platforms are increasingly relying on personalized recommender systems to help guide consumer choice. An important but understudied question in such settings is how many products to include in a recommendation set. In this work, we study how the number of recommended products influences consumers’ search and purchase behavior in an online personalized recommender system within a retargeting setting.

Methodology/Results: Via a field experiment involving 1.6 million consumers on an online retailing platform, we causally demonstrate that consumers’ likelihood of purchasing any product from the recommendation set first increases then *decreases* as the number of recommended products increases. Importantly, as much as 64% of the decrease in purchase probability (i.e., the choice overload effect) can be attributed to a decrease in consumers’ likelihood of starting a search (i.e., clicking on any recommended product). We discuss the possible behavioral mechanisms driving these results and analyze how these effects could be heterogeneous across different product categories, price ranges, and timing.

Managerial Implications: This work presents real-world experimental evidence for the choice overload effect in online retailing platforms, highlights the important role of consumer search behavior in driving this effect, and sheds light on when and how limiting the number of options in a recommender system may be beneficial to online retailers.

Key words: choice overload, retail operations, field experiment, platform operations, search cost

1. Introduction

Nowadays, with the rise of online retailing platforms, consumers are offered an ever-increasing number of options for almost every purchase decision. The plethora of options, along with technology that enables detailed tracking of consumers’ online activities, has led platforms to adopt recommender systems that proactively recommend personalized product assortments to consumers. A particularly common and effective channel for doing so is through retargeting—that is, sending recommendations to consumers who have previously interacted with products on the platform (Lambrecht and Tucker 2013, Bleier and Eisenbeiss 2015). For example, Amazon.com and Ebay.com track consumers who

*The first two authors contribute equally, and the next two authors are listed alphabetically. The last four authors are industry collaborators.

have browsed and/or used virtual shopping carts and send them recommendation messages accordingly.¹ Travel websites such as Airbnb send reminder messages (which may include recommendations) to consumers who started, but did not complete, a booking process. Spurred by the rise of generative artificial intelligence (AI), many retailing platforms are also exploring ways to include chat channels (e.g., chatbots) on their websites and use chat data for retargeting efforts.² In general, offering consumers personalized recommendations via emails or other messaging channels based on products that consumers have interacted with in the past has become a ubiquitous practice (Hosanagar et al. 2014, Schrage 2017).³

An important but understudied question in the design of such recommender systems is how many products to offer in a recommendation set. On one hand, according to standard choice theory, expanding the number of products offered should increase purchase likelihood because it increases the chance that consumers find an item that suits their preferences.⁴ The recommender system literature in operations management and computer science (Zhang et al. 2019, Feldman et al. 2022) also suggests that offering as many products as possible in recommendation results (subject to capacity constraints) is the right strategy, since each product’s expected revenue, calculated as the estimated purchase rate multiplied by the price, cannot be below zero.

On the other hand, there are reasons to believe that presenting too many products in a recommendation set may backfire and lead to less purchase. In particular, some experimental studies in psychology and marketing have documented the *choice overload* effect, whereby consumers’ purchase probability (i.e., the likelihood of making a purchase from a choice set) decreases as the choice set gets larger (for reviews, see Scheibehenne et al. 2010, Chernev et al. 2015). This effect could arise because consumers have limited capacity for processing information and hence find it more difficult to choose from a larger set (Jacoby et al. 1974, Levav et al. 2010), or because they anticipate a higher chance of making an inferior choice from a larger set and more regret from doing so (Schwartz et al. 2002).

Existing studies of the choice overload effect in retailing have mostly relied on laboratory experiments (e.g., Gourville and Soman 2005, Diehl and Poynor 2010, Hadar and Sood 2014) or small-scale offline field experiments (e.g., Iyengar and Lepper 2000), while field evidence of this effect in online

¹ <https://www.markettailor.io/blog/personalized-retargeting-ads-to-bring-customers-back-to-site>

² <https://chat.linkmobility.de/en/effective-retargeting-using-data-from-chats/>

³ Throughout this paper, “personalized recommendations” refer to recommendations based on consumers’ previous actions (e.g., browsing or inquiry activities) on the platform, rather than on consumers’ personal attributes. This definition is consistent with the personalized retargeting literature (Lambrecht and Tucker 2013, Bleier and Eisenbeiss 2015).

⁴ In standard choice theory (Luce 1959, Rieskamp et al. 2006), this intuition is formalized as the regularity property, which states that the proportion of decision makers choosing an option (e.g., the outside option) cannot increase if more options are added to the choice set.

retailing, particularly with personalized recommendations, remains scarce. It is possible that personalization or the highly specialized nature of online retailing would reduce, or even eliminate, the choice overload effect observed in the lab or offline retail settings. Hence, it remains an open question whether online retailers could benefit from limiting the number of products they recommend to consumers.

In this work, we report on a large-scale field experiment to study how choice set size influences consumer behavior with respect to the retargeting recommender system of an online retailing platform. Importantly, our empirical setting allows us to investigate not only whether consumers eventually purchase from a choice set but also whether they engage with the choice set in the first place (by clicking on any of the products in the set). The latter is an important indicator of consumers' decision-making process but is commonly missing in prior research on the choice overload effect—likely due to the difficulty of collecting data on such nuanced behavior in offline environments. In the context of recommender systems, it is crucial to understand how the choice overload effect influences consumers' click behavior for two practical reasons. First, given that click behaviors are much more frequent than purchase behaviors on online retailing platforms, click data are more useful for training algorithms to predict consumer preferences based on their past behaviors (Feldman et al. 2022). Thus, understanding drivers of click behavior—including how and why consumers' click likelihood varies with recommendation set size—is extremely important for building recommender systems. Second, if the choice overload effect exists in an online recommender system, analyzing how it applies to click versus final purchase behavior could lead to different prescriptive implications for firms. For instance, if larger recommendation sets increase consumers' tendency to click on the recommended products but reduce their purchase likelihood (e.g., because they find it difficult to make a choice after examining multiple products), the firm may be able to attenuate the latter effect by making it easier for consumers to compare the products they have clicked on, or by simplifying the presentation format of information on the product pages (Townsend and Kahn 2014, Gourville and Soman 2005). By contrast, these strategies would not be effective if larger recommendation sets lead consumers to avoid (or delay) engagement by refraining from clicking on any of the recommended products in the first place. In that case, the firm may need to consider alternative strategies such as altering the design of the initial interface that recommends products, or simply reducing the number of options recommended.

To study the impact of recommendation set size on consumers' click as well as purchase behavior, we conducted a large-scale randomized field experiment on Alibaba—the largest online retailing platform in China. Our field experiment leveraged a recommender system Alibaba developed where consumers receive recommended products in messages based on their past behavior. Our experiment ran from August to October 2020 and randomly assigned eligible consumers into one of four groups: Group 1, Group 2, Group 3, and Group 4. The four groups differed only in the number of

recommended products in the recommendation messages. Specifically, for consumers in Group n , where $n = 1, 2, 3, 4$, the recommendation message displayed n products, including one product that the consumers previously interacted with and $n - 1$ similar products. Hereafter, we refer to all products included in a recommendation message as recommended products. For each recommendation message, we tracked whether the consumer clicked on *any* recommended product in the list, as well as whether she purchased *any* recommended product through product links in the message.

Our analyses involved 1,615,325 consumers who met certain eligibility criteria (described in Section 4). Our experiment results reveal that the recommendation set size has a non-monotonic effect on consumers' purchase probability (defined as their likelihood of purchasing *any* product in the recommendation message). Specifically, compared to listing only one product in the recommendation messages, listing two products increases consumers' purchase probability by 2.05 percentage points (67% in relative effect). However, adding a third product *lowers* their purchase probability by 0.60 percentage points (12%), compared to presenting only two products in the messages. Adding a fourth product further lowers consumers' purchase probability by 0.24 percentage points (5%), compared to displaying three products in the messages. All of these differences are statistically significant (all p-values < 0.001). These results suggest that, contrary to predictions of standard choice theory but consistent with the choice overload effect, a larger recommendation set can indeed backfire and result in less consumer purchase in our setting.

Turning to consumers' click behavior, we observe that increasing the recommendation set size leads to similar non-monotonic changes in consumers' click probability (defined as their likelihood of clicking on *any* product in the recommendation message). Specifically, we find that consumers' click probability is 10.24% when they are offered only one product, increases to 20.75% when they are presented with two products, then *decreases* to 19.18% and 18.88%, respectively, when three and four products are recommended. All pairwise differences are statistically significant (all p-values < 0.001). This pattern suggests that an important driver of the choice overload effect on purchase probability in our setting could be a reduction in consumers' tendency to engage with the recommendation set (by clicking on any of the recommended products). We also show that as the recommendation set size increases, of those consumers who do click, they click on more options on average but are less likely to make a purchase. Overall, as the recommendation set size increases, our estimates suggest that as much as 64% of the reduction in purchase probability in our data can be attributed to the decrease in click probability, and the rest can be attributed to a decrease in consumers' purchase probability conditional on clicking.

We further explore how our observed effects differ across salient features of a recommendation set, including time of recommendation (i.e., weekday versus weekend), price, and product category. We show that if a recommendation message is sent on a weekday (as opposed to a weekend), the negative effect of offering a larger recommendation set on click probability is marginally significantly

stronger. This pattern may be explained by differences in search costs (i.e., the time and effort cost of clicking on a product link to inspect the product page), since consumers may have higher search costs on weekdays than weekends. We also show that the negative effect of recommendation set size on click probability is marginally significantly more prominent for more expensive products and is not the same across different categories. Last but not least, we apply the causal forest method to identify the combinations of recommendation features under which increasing recommendation set size is most (or least) likely to discourage consumer search.

In summary, our research provides, as far as we know, the first empirical evidence from a large-scale field experiment that both click and purchase behavior could decrease with the size of the recommendation set in an online retail setting. These results contribute to the literature on the choice overload effect and shed light on how this effect could influence business profits for online retailers. Moreover, our findings have important practical implications for the optimal design of recommendation messages in retailers’ retargeting campaigns. In particular, our findings shed light on the conditions under which limiting the number of recommended products could be more conducive to consumer engagement and purchase.

The rest of this paper is organized as follows. We review the relevant literature in Section 2. In Section 3, we discuss the possible behavioral mechanisms and generate the main hypotheses. Section 4 describes the field experiment setting. The main results of the field experiment are presented in Section 5, and we discuss heterogeneous treatment effects in Section 6. Section 7 concludes with a discussion of managerial implications, robustness checks, and future research directions.

2. Literature Review

Our research is closely related to several broad streams of literature: (i) empirical research in retail operations and platform algorithm design, (ii) retargeting strategies, and (iii) consumer behavior in online retailing contexts.

2.1. Empirical Retail Operations and Platform Algorithm Design

First, our work contributes to the empirical literature on retail operations that investigates how retailers’ or retail platforms’ operational practices could shape consumer engagement and purchase behavior. In brick-and-mortar retail settings, studies have documented how consumers’ purchase behavior can be affected by inventory availability (Musalem et al. 2010, Cachon et al. 2019, Li et al. 2023), waiting in queues (Lu et al. 2013), adoption of new technologies (Stamatopoulos et al. 2021), and operational transparency (Buell and Kalkanci 2021), among other operational factors. In online retail contexts, researchers have also empirically investigated the impact of various operational levers, such as pricing (Moon et al. 2018, Zhang et al. 2020, Lu et al. 2022), information provision (Gallino and Moreno 2018, Cui et al. 2019, Calvo et al. 2023, Hu and Wang 2022), design of search processes (Ngwe et al. 2019), and website speed (Gallino et al. 2023). In particular, a number of these studies

find that operational practices could influence consumers’ purchase behavior through shaping their search behavior. For example, Cachon et al. (2019) find that displaying more inventory could reduce sales because more inventory at one store induces consumers to postpone decision-making and engage in more search at other stores. Ngwe et al. (2019) show that increasing search frictions (e.g., by removing links to discount sites or removing the option to rank search results by discount percentage) could help improve online retail firms’ profits by inducing consumers to search more products. Zhang et al. (2020) demonstrate that shopping cart promotions could boost customer engagement in the long run by increasing their product views and purchases on a retail platform. Our work contributes to this literature by identifying a new important lever—the recommendation set size—for influencing consumers’ search and purchase behavior in online retail.

In addition, our paper relates to the literature that designs artificial intelligence algorithms for online platforms and tests them via field experiments—in a variety of contexts, from dynamic pricing (Ferreira et al. 2016, Fisher et al. 2018) and online image display (Khern-am nuai et al. 2024) to product recommendation systems (Feldman et al. 2022) and matching (Rios et al. 2022). In particular, some of these papers have demonstrated the value of understanding and properly modeling consumer choice behavior in the design of recommendation algorithms. For example, Feldman et al. (2022) show that a featurized multinomial logit (MNL) choice model applied to a recommender system on Alibaba’s online marketplaces could significantly improve revenue per customer visit, compared to the traditional machine-learning algorithm that does not explicitly model consumer choice behavior. Rios et al. (2022) study the optimal design of assortment sets to maximize match rates in a dating app and show that their algorithm, which incorporates customers’ sequential interaction behavior on the platform into the choice model, yields more matches relative to their industry partner’s algorithm. We contribute to this literature by quantifying the impact of recommendation set size on consumer search and purchase behavior, which could in turn help inform the optimal design of recommendation algorithms.

2.2. Retargeting

Our work also contributes to a limited but growing body of research that studies retailers’ retargeting strategies. Broadly defined, retargeting refers to the practice of reaching out (via advertisement banners, email, in-app message, SMS text, etc.) to consumers who have signalled their interest in the retailer’s products in some way, including browsing the website (Lambrecht and Tucker 2013, Sahni et al. 2019), creating a user account (Hoban and Bucklin 2015), or adding products to cart (Li et al. 2021). Previous work in this area has focused on investigating the effectiveness of retargeting messages (Hoban and Bucklin 2015, Johnson et al. 2017, Sahni et al. 2019), the value of personalized recommendations (i.e., recommendations based on consumers’ most recent product engagement behavior) in these messages (Lambrecht and Tucker 2013, Bleier and Eisenbeiss 2015), and the optimal times at which they should be sent (Bleier and Eisenbeiss 2015, Sahni et al. 2019, Li et al. 2021).

By contrast, we take personalization and timing as given and study how many products to recommend in a personalized retargeting message. The empirical setting that we study (see more details in Section 4) is one where personalized retargeting may be particularly effective, based on insights from extant research. That is, previous research has found personalized retargeting to be most effective when (i) consumers have established narrowly construed preferences (which usually happens later in the purchasing funnel; Lambrecht and Tucker 2013), and (ii) the retargeting messages are sent soon after consumers engage with a product (Bleier and Eisenbeiss 2015). Our empirical setting possesses those two features; as such, our novel insights on recommendation set size as a potential moderator for retargeting effectiveness could be particularly relevant and valuable for practice.

2.3. Consumer Search and the Choice Overload Effect

By investigating consumers' click behavior online, our work relates to the literature on consumer search (Stigler 1961). In particular, there is a large body of work on empirically estimating search cost with both purchasing data and (increasingly) online click data (Kim et al. 2010, De los Santos et al. 2012, Honka 2014, Chen and Yao 2017, Jiang et al. 2021). A number of these papers have also applied the estimates from their structural models to run counterfactual studies that quantify the value of online search-enhancing tools, such as product recommendations (Kim et al. 2010), search refinement (Chen and Yao 2017) and retargeting advertisements (Jiang et al. 2021). However, the aforementioned works all adopt the perspective of standard choice models and assume that the marginal search cost does not depend on the number of options available. Hence, their models preclude the possibility that offering consumers more options could lower consumers' propensity to start a search (or make a purchase). From this perspective, we add to this literature by highlighting that consumer search could be deterred not only by the time and effort traditionally assumed in marginal search costs, but also a *decision* cost that could increase with the number of options.

Relatedly, a number of papers have studied position (ranking) effects in search results (Ghose et al. 2012, De los Santos and Koulayev 2017, Ursu 2018). For instance, Ursu (2018) use randomized field experiments to show that, while higher position may be associated with lower search cost, it does not affect consumers' expectation of utility or realized value after search. Researchers in operations management have also studied assortment optimization policies when consumers do not consider all available options due to search costs (Wang and Sahin 2018, Gallego et al. 2020). Again, these papers do not consider the possibility that the marginal search cost for any given position changes with the number of options, and hence would not predict a lower search propensity if the choice set increases.

Last but not least, our work adds to the literature on the choice overload effect (which we review in more detail in the next section). This literature stems from the seminal work of Iyengar and Lepper (2000), who found via a field experiment that consumers at a grocery store were less likely to purchase a jar of jam on days when 24 flavors (versus 6 flavors) were displayed. They further

conducted two laboratory experiments that identified a similar effect in the context of essay-writing and chocolate purchasing. Follow-up studies of this effect, however, have yielded mixed evidence (Chernev et al. 2015), leading some researchers to question its existence (Simonsohn et al. 2014) and others to conjecture that the existing laboratory experiments are severely underpowered (Dean et al. 2022). Our paper extends this stream of work by examining the choice overload effect in a new and highly relevant context (i.e., personalized recommender system of an online retail platform) with a large-scale field experiment. Further, while Iyengar and Lepper (2000) provide suggestive evidence that a larger choice set may attract more attention from consumers but lead to less purchase⁵, we employ a more precise measurement of engagement (i.e., clicking data) and show that in our empirical setting, a larger recommendation set could *reduce* consumers’ propensity to engage with the set. We also note that while a large amount of work in computer science and information systems has shown the effectiveness of recommender systems in general (e.g., see Lü et al. 2012, Bobadilla et al. 2013 and Zhang et al. 2019 for extended reviews of recommender systems), little is known about the effect of recommendation set size on consumer behavior.⁶ Thus, our work contributes to this literature by demonstrating the importance of recommendation set size on the effectiveness of recommender systems.

3. Hypotheses Development

In this section, we discuss the possible mechanisms driving the impact of recommendation set size on consumers’ search and purchase decisions based on the extant literature, and we derive our hypotheses.

As mentioned earlier, standard choice theory suggests that a larger recommendation set should always increase consumers’ likelihood of starting a search (i.e., clicking on any of the products in the recommendation set) and making a purchase (i.e., buying any of the recommended products; see, e.g., Talluri and Van Ryzin 2004). This is because, with more options, consumers should expect a higher likelihood of finding an option that suits their preferences, which in turn should increase their willingness to engage with the recommendation set by clicking on at least one of the recommended products. To formalize this intuition in our context, we present and analyze a stylized two-stage search-and-choose model (see Appendix C.1). We assume that consumers incur a search cost for each product they click on, which captures the time and effort it takes the consumer to collect product information (e.g., inspecting detailed product descriptions, watching use videos, reading reviews) as

⁵ They recruited volunteers to count the percentage of people who approached the jam display table when passing by.

⁶ The only exception is Bollen et al. (2010), who use scenario-based experiments to measure the number of movie recommendations (5 or 20) on subjects’ perceptions. They propose that the increased attractiveness of larger sets is offset by increased decision difficulty, resulting in no significant difference in choice.

well as to deliberate whether the product matches the consumer’s individual needs.⁷ We prove that, under standard assumptions of consumer choice, including more products in the recommendation set should always lead to a (weakly) higher likelihood of clicking on any of the products, as well as a (weakly) higher likelihood of making a purchase.

Next, based on existing research on the choice overload effect (e.g., Scheibehenne et al. 2010, Chernev et al. 2015), we discuss two main behavioral mechanisms under which a larger recommendation set size may reduce consumers’ purchase probability. We also highlight how these two mechanisms diverge in their predictions on consumers’ search behavior. We note that these mechanisms may not be the *only* mechanisms that could play a role in driving consumers’ behavior in our empirical setting, but they are well-documented in the prior literature and useful in illustrating the possible predictions. We discuss additional possible mechanisms in Section 5.3.

3.1. Anticipated Regret

One possible mechanism for the choice overload effect is anticipated regret (Schwartz et al. 2002, Irons and Hepburn 2007, Sarver 2008). According to regret theory, emotions of regret arise when people make decisions under uncertainty and, after the uncertainties resolve, realize they should have chosen a different option that is ex post superior (Loomes and Sugden 1982, Zeelenberg 1999). Intuitively, in larger choice sets, the best unchosen option is likely to be more attractive, which then leads to a higher likelihood of consumers experiencing regret and hence feeling dissatisfaction with their choice. Indeed, a number of laboratory studies have found that subjects express more postdecision regret or lower satisfaction with their choice when choosing from a larger (vs. smaller) set (Iyengar and Lepper 2000, Sagi and Friedland 2007, Reutskaja and Hogarth 2009, Inbar et al. 2011). Anticipating the negative consequences of making an inferior choice and recognizing the greater likelihood of picking an inferior option from a larger set, consumers may become less likely to make a choice from a larger set. Gourville and Soman (2005) directly test and confirm the role of anticipated regret as a driver of the choice overload effect in the lab: they show that the choice overload effect is diminished when postpurchase exchanges are allowed (which should reduce potential regret since decisions can be altered after uncertainties resolve).

Applied to our empirical context—which includes both the search and the purchase stages—the anticipated-regret mechanism suggests that when presented with a larger recommendation set, consumers would first feel compelled to search (i.e., click on) more of the recommended products. This is because if the number of searched products is the same, consumers would anticipate more regret from choosing an ex post inferior product when the recommendation set is larger. Since

⁷ Such a search cost may apply even to a product that the consumer has previously interacted with. Our model predictions remain robust if (1) we assume a very low search cost (or very low value uncertainty) for the first product in a retargeting recommendation message (e.g., because the consumer has previously learned about the first product) and/or (2) we include additional behavioral effects of retargeting recommendation messages (e.g., their effect as a reminder of previously viewed products); see Appendix C for more discussion.

search is costly, consumers may then refrain from (or delay) starting a search in the first place. Although prior theoretical studies have modeled anticipated regret as an explanation of the choice overload effect (Irons and Hepburn 2007, Sarver 2008), their models do not separately account for consumers' search and purchase behavior. Therefore, to formalize the intuition and generate testable predictions, we formally model this mechanism (in combination with standard choice theory) in our two-stage setting (see Appendix C.2), and we show that as the number of choices increases, both search and purchase likelihoods first increase then decrease.⁸ Under this mechanism, the choice overload effect is essentially caused by an additional *decision* cost that is a function of (1) consumers' anticipated regret from making a purchase decision without searching all available options and (2) search costs, and such a decision cost increases with the number of products in the recommendation set. Importantly, our theoretical analyses show that under this mechanism, the negative effects of a larger recommendation set become weaker when consumer search cost or anticipated regret is lower (see Corollary 1 and Figure A8 in Appendix C.2).

3.2. Information Overload

Another reason that offering too many options may deter purchase is that the decision maker may be overwhelmed by information relating to the choices (Jacoby et al. 1974, Timmermans 1993, Pocheptsova et al. 2009). Laboratory studies have shown that subjects tend to experience greater decision difficulty and report greater frustration with the decision-making process when choosing from a larger choice set (Iyengar and Lepper 2000, Lee and Lee 2004). Gourville and Soman (2005) further hypothesize that if the choice overload effect is driven by the cognitive difficulty of processing information, then reducing the cognitive load by simplifying the way information is presented could alleviate this effect. Indeed, consistent with this proposition, laboratory studies have shown that when subjects are presented with a simplified table of information (rather than the full version; Gourville and Soman 2005) or fewer attributes of the available options (Greifeneder et al. 2010), the choice overload effect is alleviated. Moreover, building on the notion that processing larger choice sets depletes mental resources faster, Levav et al. (2010) hypothesize and show empirically that making decisions about larger choice sets early on renders consumers more likely to choose the default option for subsequent decisions.

⁸ An alternative mechanism that has been proposed to explain the choice overload effect is *strategic inference* (Kuksov and Villas-Boas 2010), whereby consumers infer from the size of the choice set the locations of the included products in the preference space (which in turn determine how long it would take them to find themselves a good option). The model in Kuksov and Villas-Boas (2010) does not match the recommendation setting we empirically study, because the model requires that the firm offers different products for each choice set size (i.e., the products don't overlap), whereas in our case, products in the smaller recommendation set are a subset of products in the larger set. Nevertheless, in our empirical setting, it is plausible that consumers believe (correctly or not) that the recommended products in a larger set are overall less tailored to their specific tastes, and hence become less likely to engage with the recommendation. A model with this mechanism would lead to similar predictions as those from a model incorporating anticipated regret.

It is generally assumed that people do not anticipate the depletion of their mental resources and therefore do not optimally allocate these resources in advance (see, e.g., Levav et al. 2010). Thus, while information overload might render consumers less likely to make a purchase after starting their search, this mechanism should not influence consumers’ propensity to search in the first place (i.e., *before* they are exposed to detailed information). In our two-stage model (see Appendix C.1), this mechanism can be operationalized as the consumer having a larger outside option value in stage 2 when the number of searched options increases. This would result in a negative relationship between the number of options a consumer searches and her likelihood of making a purchase in stage 2, but would not change the consumer’s search decision in stage 1. Therefore, we expect that if the information overload mechanism is the dominant driver of consumer response to a large choice set, then as the recommendation set size increases, consumers will become more likely to start a search (as predicted by standard choice theory), but may be less likely to make a purchase given that they search.

3.3. Hypotheses

To summarize, while standard choice theory predicts that consumers’ purchase probability should always (weakly) increase in recommendation set size, behavioral mechanisms such as anticipated regret and information overload predict the choice overload effect, whereby the purchase probability could decrease as recommendation set size increases. Combining standard choice theory with these behavioral mechanisms, we hypothesize the following:

HYPOTHESIS 1. *As recommendation set size increases, consumers’ likelihood of making a purchase first increases then decreases.*

However, the two behavioral mechanisms—anticipated regret and information overload—diverge in their predictions about consumers’ search behavior. Specifically, combining the two behavioral mechanisms with standard choice theory, we derive the following competing predictions.

HYPOTHESIS 2A. *As recommendation set size increases, consumers’ likelihood of starting a search (i.e., clicking on any product) first increases then decreases.*

HYPOTHESIS 2B. *As recommendation set size increases, consumers’ likelihood of starting a search (i.e., clicking on any product) increases.*

In the next section, we present a field experiment to test these hypotheses.

4. Randomized Field Experiment Overview and Data

We conducted a large-scale field experiment at Alibaba Group, a leading e-commerce company in China, to explore the impact of recommendation set size on consumer choice behavior. In this section, we introduce our field setting and experimental design, followed by an overview of our experimental data and the randomization check results.

4.1. Field Setting

The field experiment was run on two online retail marketplaces of Alibaba Group: Tmall and Taobao. Tmall is the largest business-to-consumer online retailing platform for brand-name goods in China, while Taobao is the largest consumer-to-consumer retailing platform for small businesses and individual entrepreneurs in China. In fiscal year 2022, the total gross merchandise volume (GMV) transacted on Alibaba’s retail marketplaces in China reached RMB 8,317 billion (USD 1,312 billion), consisting mostly of transactions on Tmall and Taobao.⁹ For simplicity, we refer to Tmall and Taobao jointly as the *Alibaba platform*.

The Alibaba platform offers an instant customer-service messenger through which consumers and merchants can communicate with each other. For example, consumers can inquire about product details or send special requests to merchants via the messenger before they make purchase decisions. Customer service representatives from the merchants will reply to consumers’ questions and requests. According to the platform, approximately 35% of consumers make an inquiry with a merchant before purchasing from the merchant. Merchants can also recommend products to consumers through the messenger.

Prior to our field experiment, Alibaba implemented an automated messaging service that allows merchants to push product recommendations through the instant customer-service messenger. Specifically, if a consumer contacts a merchant about a product but does not order that product within five minutes of the inquiry, the messenger will automatically send a recommendation from the merchant to the consumer. The recommendation message displays one or more products sold by the merchant and selected by Alibaba’s recommendation algorithm, including the product the customer recently asked the merchant about. The recommendation message has the same design across merchants: It displays products in “tabs”, with each tab showing the name, price, and picture of one recommended product and linking the consumer to the corresponding product page. In addition, merchants have the discretion to add some text to encourage consumers to consider the products included in the recommendation message (referred to hereafter as “recommended products”), and the additional text content is determined by merchants. Figure A1a in Appendix A shows an example of a post-inquiry recommendation.

Upon receiving a post-inquiry recommendation, the consumer will be notified in the app: a red dot will appear on the message-box icon, as shown in Figure A1b in Appendix A.¹⁰ The consumer can click on the message box icon to view the recommendation message, which will appear below their most recent inquiry with the merchant. The consumer can then click on the product tabs in the recommendation message to visit the corresponding product pages, review product information, and place an order. Figure A2 in Appendix A summarizes the flow of a consumer’s experiences and actions regarding the recommendation.

⁹ <https://data.alibabagroup.com/ecms-files/1532295521/34596e84-ff0c-4522-bbc8-82a8c49c0bba.pdf>.

¹⁰ If Alibaba’s app is running and its notification service is enabled on the consumer’s phone, the consumer will also receive a push notification in the operating system about the recommendation message.

4.2. Experimental Design

Our field experiment leveraged this automated product-recommendation service. We next describe how recommended products were selected and ranked in recommendation messages in our experiment. In general, each message contained a list of n recommended products, where n was a specified integer depending on the experimental group. The n recommended products consisted of one inquired product and $n - 1$ products that were sold by the same merchant and deemed similar to the inquired product by Alibaba’s algorithm (hereafter, “similar products”).¹¹ The inquired product was always displayed first in the recommendation message, followed by the $n - 1$ similar products displayed in descending order of their similarity to the inquired product, where the similarity was calculated by Alibaba’s proprietary algorithm based on product attributes. Our context resembles typical product recommendation systems in online retail, where the recommended products are ranked based on the extent to which they are expected to match consumers’ interests (Feldman et al. 2022). In particular, we may reasonably infer that consumers are more interested in the inquired product than in products they have not inquired about, and thus similarity to the inquired product may reflect how closely other products on average match consumers’ preferences.¹² Our setting also reflects a common practice in designing recommender systems, where personalized recommendations are chosen based on products consumers have previously engaged with (e.g., Li et al. 2021) and often include other products deemed similar by retailers.

If a consumer inquired about more than one product, only the inquired product with the highest predicted purchase probability would be shown in the recommendation message.¹³ In the rest of the paper, “inquired product” specifically refers to the inquired product that was displayed in the recommendation message. If there were more than $n - 1$ similar products for the inquired product, the recommendation message would display the $n - 1$ products with the greatest similarity to the inquired product. Notably, for a given inquired product, the selection and ranking of similar products depended only on product attributes and were not affected by any consumer-specific attributes. Therefore, the list of products in a recommendation message was uniquely determined by the inquired product and the number of recommended products n , regardless of the consumer making the inquiry.

We conducted the field experiment from August 15, 2020 to October 14, 2020. The experiment involved a randomly selected subset of consumers who met the following criteria: (1) they contacted a merchant via the customer-service messenger about at least one product during the experimental

¹¹ Similar products in this setting represent different products (e.g., jeans in different styles) rather than the same product with differences in minor features (e.g., the same jean in different colors or sizes).

¹² Our data confirm this conjecture. Specifically, our experimental results show that the click rate and purchase rate of the inquired product are higher than those of any other recommended product in the message; see Table 3 and the discussion in Section 5 for more details.

¹³ The majority (94%) of consumers in our sample inquired about only one product before receiving the recommendation message included in our analysis.

period and were eligible to receive a recommendation message, and (2) their inquired product had at least three similar products as defined by Alibaba’s algorithm. The selected consumers were then randomly assigned into one of four experimental groups: Group 1, Group 2, Group 3, and Group 4. Figure A3 in Appendix A outlines the different recommendation messages sent to the four treatment groups. For consumers in Group n ($n = 1, 2, 3, 4$), they received a message recommending n products, with the inquired product at the top followed by the $n - 1$ similar products. Note that in our experiment, a recommendation message could contain at most three similar products, in addition to one inquired product. Since our experiment targeted only consumers whose inquired product had at least three similar products, every consumer in our experiment could potentially be recommended four products. Our subjects are thus comparable across treatment groups after randomization. The procedures to select and rank the recommended products (as introduced earlier) were the same for all groups. If a merchant decided to include additional text to be automatically sent along with recommendation messages, the text was preset by the merchant and did not vary with the number of recommended products in the messages. Therefore, the four experimental groups differed only in the number of recommended products in the messages. The recommendation messages were designed to fit entirely on mobile screens, with all content in the recommendation message visible without scrolling, even when there were four recommended products.

4.3. Data and Randomization Check

During the two-month experimental period, we observed 3,827,877 recommendation messages sent to 2,366,182 consumers who satisfied our aforementioned criteria. For each recommendation message, we tracked the consumer’s click and purchase behavior through the recommendation channel on the day of receiving the message.¹⁴ To ensure an accurate and clean estimate of how the size of a recommendation set affects search and purchase behavior, we consider only consumers who viewed at least one recommendation message, and we analyze only the first recommendation message each consumer viewed during the experiment. We focus on consumers who viewed at least one message, because only by viewing a message could they see the recommendation set and thereby become exposed to the experimental treatment (i.e., the number of recommended products). Importantly, the number of recommended products in a message could not have influenced consumers in any way before they viewed the first recommendation message; thus, our experimental design should guarantee that random assignment was maintained across groups among consumers who viewed at least one message. We confirm this balance in our randomization check, as reported in the next paragraph.

¹⁴ Tracking consumers’ engagement with marketing messages (e.g., promotional messages or recommendation messages) during the day of receiving a message is a common practice at Alibaba. In our specific setting, purchases made on the day of receiving a recommendation message account for more than 90% of all purchases made via the message channel within a week. Therefore, click and purchase behavior on the day of receiving a recommendation message was deemed sufficient by the company for comparing the performance of different experimental groups, and this is why Alibaba shared data only within this time window with us.

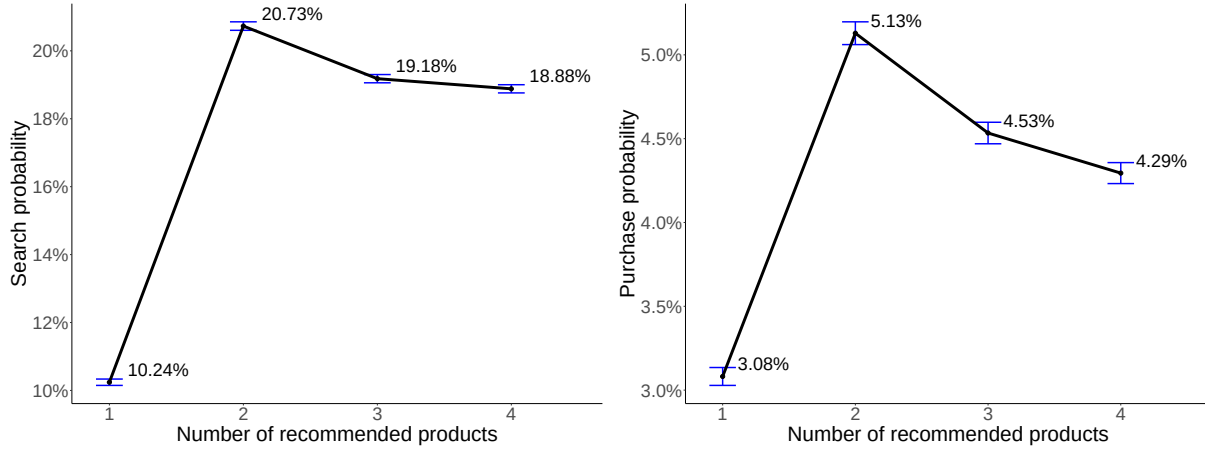
We focus only on the first recommendation message viewed by a consumer, because viewing the first message may change consumers’ future inquiry behavior, which could affect the existence, timing, and content of subsequent recommendation messages and in turn bias our estimates had we included subsequent messages in the analyses. Nevertheless, our empirical results are robust if we analyze all sent messages or all viewed messages.

Our final analysis sample consists of 1,615,325 messages sent to 1,615,325 consumers. The number of messages in each group (which is the same as the number of consumers in each group, since our analysis includes one message per consumer) is shown in Table A1 in Appendix A. To confirm that the inquired products and recommendation messages in our final analysis sample were similar across the four experimental groups, Table A1 compares across groups two important characteristics made available to us—the log-transformed price of inquired products and whether a message was sent on weekends (as opposed to on weekdays).¹⁵ For all groups, the average log-transformed price of inquired products in Chinese Yuan (CNY) was 4.87, and the percentage of messages sent on weekends was approximately 31.8%. We performed pairwise t-tests for the log-transformed price of inquired products and pairwise two-sample proportion tests for the proportion of messages sent on weekends. The four groups were statistically indistinguishable in these metrics, as the minimum p-value across all pairwise statistical tests is greater than 0.17. This randomization check suggests that the messages received by consumers in each group were statistically similar.

5. Field Experiment Results

In this section, we present the main results of our field experiment. Our unit of analysis is a recommendation message sent to consumer i from merchant j on day t . Because we consider only the first message viewed by each consumer, the unit of analysis is also at the consumer level. To test our hypotheses in Section 3 pertaining to how the size of a recommendation set affects engagement with and purchase from the choice set, we construct two outcome variables to capture whether consumers searched any recommended product as well as whether consumers purchased any recommended product directly via the recommendation message, on the day of receiving the message. Specifically, $SearchIncidence_{ijt}$ is a binary variable that equals one if consumer i clicked on *at least one* recommended product in the recommendation message sent by merchant j on day t and zero otherwise; $PurchaseIncidence_{ijt}$ is a binary variable that equals one if consumer i purchased *at least one* recommended product directly from the product links in the recommendation message sent by merchant j on day t and zero otherwise. Note that if $SearchIncidence_{ijt} = 0$, then $PurchaseIncidence_{ijt} = 0$, since we are interested only in purchases made via the recommendation message (as opposed to other channels).

¹⁵ Due to Alibaba’s privacy concerns, we did not receive data on consumer attributes and thus could not conduct randomization checks on these attributes.



Note: The error bars depict the 95% confidence intervals for the search probability and purchase probability.

Figure 1 Search Probability (Left) and Purchase Probability (Right) as a Function of Recommendation Set Size

Table 1 Summary Statistics and Group Comparisons of Consumer Search and Purchase Behavior

	Group 1 (1 Product)	Group 2 (2 Products)	Group 3 (3 Products)	Group 4 (4 Products)
Search Probability	10.24%	20.73%	19.18%	18.88%
Marginal Difference	-	10.49%	-1.55%	-0.30%
(P-value)	-	(< 0.001)	(< 0.001)	(< 0.001)
Purchase Probability	3.08%	5.13%	4.53%	4.29%
Marginal Difference	-	2.05%	-0.60%	-0.24%
(P-value)	-	(< 0.001)	(< 0.001)	(< 0.001)

Notes: This table compares search and purchase probabilities across groups. “Marginal difference” represents the differences in search probabilities (or purchase probabilities) between consecutive columns. We report p-values of two-sample proportion tests for all comparisons in this table.

5.1. Main Results

Figure 1 and Table 1 display consumers’ search and purchase probabilities, which equal the mean values of $SearchIncidence_{ijt}$ and $PurchaseIncidence_{ijt}$, respectively, for each group. Both the search and purchase probabilities exhibit an inverted-U-shaped pattern: they first increase then decrease as the number of recommended products increases. Compared to recommending only one product (where search and purchase probabilities are 10.24 percentage points and 3.08 percentage points, respectively), recommending two products increases the search probability by 10.49 percentage points (102%) and boosts the purchase probability by 2.05 percentage points (67%). However, adding a third product lowers the search probability by 1.55 percentage points (7%) and purchase probability by 0.60 percentage points (12%), relative to recommending two products. Adding a fourth product further reduces the search probability by 0.30 percentage points (2%) and purchase probability by 0.24 percentage points (5%), relative to recommending three products. All the differences reported above are statistically significant (all p-values < 0.001 for two-sample proportion tests).

We next conduct regression analyses to examine whether the preceding effects are robust to controlling for other variables. Specifically, we use the following ordinary least squares (OLS) regression

Table 2 OLS Regressions Predicting Consumer Search and Purchase Behavior

Outcome Variable	SearchIncidence		PurchaseIncidence	
	Baseline: Group 2 (1)	Baseline: Group 3 (2)	Baseline: Group 2 (3)	Baseline: Group 3 (4)
Group 1	−0.1044**** (0.0011)	−0.0887**** (0.0010)	−0.0204**** (0.0005)	−0.0143**** (0.0005)
Group 2		0.0156**** (0.0009)		0.0061**** (0.0005)
Group 3	−0.0156**** (0.0009)		−0.0061**** (0.0005)	
Group 4	−0.0181**** (0.0009)	−0.0024*** (0.0009)	−0.0082**** (0.0005)	−0.0021**** (0.0005)
Merchant Fixed Effects	Yes	Yes	Yes	Yes
Date Fixed Effects	Yes	Yes	Yes	Yes
Observations	1,615,325	1,615,325	1,615,325	1,615,325

Notes: Columns (1) and (2) report the estimated differences in search probabilities between different groups, with Group 2 and Group 3 as the baseline group, respectively. Columns (3) and (4) report the estimated differences in purchase probabilities between different groups with Group 2 and Group 3 as the baseline group, respectively. Standard errors are robust and clustered at the merchant level. *p<0.1; **p<0.05; ***p<0.01; ****p<0.001.

model:

$$OutcomeVariable_{ijt} = \alpha_0 + \sum_{n \in N} \alpha_1^n Group_i^n + X_j + D_t + \epsilon_{ijt}. \quad (1)$$

In Specification (1), $OutcomeVariable_{ijt} \in \{SearchIncidence_{ijt}, PurchaseIncidence_{ijt}\}$, $Group_i^n$ is a binary variable that equals 1 if consumer i was assigned to Group n and 0 otherwise, X_j and D_t represent merchant and date fixed effects, respectively, and $N = \{1, 2, 3, 4\} \setminus \{n^*\}$, where Group n^* is the baseline group. To estimate the marginal effect of adding recommended products, we run the regression twice for each outcome variable with the baseline group being Group 2 ($N = \{1, 3, 4\}$) and Group 3 ($N = \{1, 2, 4\}$), respectively. We report the robust standard errors clustered at the merchant level in all regression analyses in this paper. Our results are robust if we use logistic regressions to predict our binary outcome variables.

Table 2 presents the results of Specification (1). The coefficients on the group indicators represent the differences in a given outcome variable between the baseline group and each of the other three groups. Compared with results reported in Table 1, our OLS regressions yield qualitatively consistent results and, in particular, demonstrate the inverted-U-shaped pattern of search and purchase probabilities as the number of recommended products increases. Specifically, columns (1) and (3) show that adding the second product increases the search probability by 10.44 percentage points and the purchase probability by 2.04 percentage points, compared to having only one product in a recommendation message. However, adding the third product lowers the search probability by 1.56 percentage points and the purchase probability by 0.61 percentage points, compared to recommending two products. Adding the fourth product further lowers search probability by 0.24 percentage points and purchase probability by 0.21 percentage points, compared to recommending three products, as shown in columns (2) and (4).

Furthermore, we can compute the extent to which the drop in search probability drives the drop in purchase probability. Specifically, Table 1 shows that when the number of products in the recommendation set increases from two to three, 64% of the drop¹⁶ in purchase probability (from 5.13% to 4.53%) can be attributed to the drop in search probability (from 20.73% to 19.18%), and the remaining 36% is driven by the drop in purchase probability conditional on search. Similarly, 30% of the drop¹⁷ in purchase probability from recommending three products to recommending four products (i.e., 4.53% to 4.29%) can be credited to the drop in search probability (from 19.18% to 18.88%). Altogether, the reduction in search probability drives 30% to 64% of the drop in purchase probability brought by the choice overload effect in our data.

Overall, we identify evidence of the choice overload effect in a recommender system, whereby a larger recommendation set could lead to a lower purchase probability. Moreover, we show that this effect is strongly driven by a reduction in consumers' likelihood of starting a search. Thus, our results support Hypotheses 1 and 2A, while rejecting Hypothesis 2B. Next, in Section 5.2, we provide several additional observations from the experimental data. We then discuss additional possible mechanisms in Section 5.3.

5.2. Additional Observations

To obtain further insights into consumers' behavior, we go beyond search and purchase probabilities at the recommendation set level (i.e., whether people search or purchase *any* product in the recommendation set) to investigate the click rate (i.e., probability of clicking) and purchase rate (i.e., probability of purchasing) for *each product* or *combination of products* in the recommendation set.

We first observe that in each group with more than one recommended product, both click and purchase rates are highest for the first recommended product, second highest for the second recommended product, and so on. Table 3 illustrates this pattern in Group 4. This observation implies that, on average, the products recommended decrease in their attractiveness to the consumer, which is consistent with the underlying rationale for the design of the recommendation messages (as described in Section 4).

Next, recall from our discussions in Section 3.1 that based on the mechanism of anticipated regret, when presented with a larger recommendation set, consumers may become less likely to start a search, but those who do start a search would search more options. Our data confirm this prediction (see Table 4).

Further, from the statistics in Table 1, we can observe that consumers' purchase probability conditional on starting a search (hereafter, "conditional purchase probability") decreases with recommendation set size (i.e., 30.08%, 24.75%, 23.62%, and 22.72% from Group 1 to Group 4, respectively).

¹⁶ $5.13\% \times (1 - 19.18\%/20.73\%)/(5.13\% - 4.53\%) = 64\%$.

¹⁷ $4.53\% \times (1 - 18.88\%/19.18\%)/(4.53\% - 4.29\%) = 30\%$.

Table 3 Comparison of Click and Purchase Rates of Each Recommended Product in Group 4

	Product 1	Product 2	Product 3	Product 4
Click Rate	13.34%	6.18%	5.85%	5.66%
Marginal Difference	-	-7.16%	-0.33%	-0.19%
(P-value)	-	(< 0.001)	(< 0.001)	(< 0.001)
Purchase Rate	3.80%	0.28%	0.18%	0.15%
Marginal Difference	-	-3.52%	-0.10%	-0.03%
(P-value)	-	(< 0.001)	(< 0.001)	(< 0.001)

Notes: This table compares the click and purchase rates of each recommended product in Group 4. Product 1, 2, 3, and 4 refer to the 1st, 2nd, 3rd, and 4th recommended product, respectively. “Marginal difference” represents the differences in click rates (or purchase rates) between consecutive columns. We report p-values of McNemar’s test for all product comparisons in the table.

Table 4 Comparison of Average Number of Clicked Products Among Consumers Who Clicked

	Group 1	Group 2	Group 3	Group 4
Average Number of Clicked Products	1.00	1.29	1.51	1.64
Marginal Difference	-	0.29	0.22	0.14
(P-value)	-	(< 0.001)	(< 0.001)	(< 0.001)

Notes: This table compares across groups the average number of clicked products among consumers who clicked. “Marginal difference” represents the differences in the average number of clicked products between consecutive columns. We report p-values of Wilcoxon rank sum tests for all group comparisons in this table.

This pattern could potentially be explained by either information overload (i.e., consumers search more and hence become less likely to make a decision as the recommendation set grows larger) or anticipated regret (i.e., consumers become less likely to make a decision when there are unsearched options). Notably, the decrease in conditional purchase probability as the recommendation set grows is not driven by consumers who search more options—contrary to the prediction of the information overload mechanism. For instance, consumers who click on three products in Group 3 are on average not less likely—in fact, they are directionally more likely—to make a purchase than consumers who click on two products in Group 2 (30.79 percentage points vs. 30.43 percentage points; p-value = 0.48). Instead, the decrease in conditional purchase probability appears to be driven primarily by consumers searching the same number of products but exhibiting a lower purchase likelihood when the recommendation set size is larger. For example, among consumers who click on two products, conditional purchase probability is lower in Group 3 than in Group 2 (22.64 percentage points versus 30.43 percentage points, p-value < 0.001). This pattern is consistent with the mechanism of anticipated regret, which predicts that consumers are less likely to purchase when there are unsearched options, and unsearched options occur more often in larger choice sets.

Finally, a question of potential interest is whether the addition of a lower-ranked item in a choice set increases consumers’ likelihood of choosing a higher-ranked one. This phenomenon is referred to as the decoy effect and has been identified in several lab and field experiments (see, e.g., Wu and Cosguner 2020 and the references therein). Our empirical setting is not designed to cleanly test the decoy effect (as defined by prior research) because to do so, the lower-ranked item should be *inferior* to the higher-ranked item for all consumers, and consumers should be able to make only one choice.

Neither of these conditions holds in our empirical setting due to consumers’ heterogeneous preferences and the possibility that they search or purchase multiple products. Nevertheless, we report several interesting patterns in our data for exploratory purposes. First, when the recommendation set size increases from one to two, we find that both the click and the purchase rates for the first product (i.e., the inquired product) increase (from 10.24 percentage points to 16.72 percentage points for click rate, and from 3.08 percentage points to 4.74 percentage points for purchase rate; both p -values < 0.001), while the conditional purchase rate for product 1 (i.e., the likelihood of purchasing the inquired product conditional on searching it) decreases (from 30.09 percentage points to 28.35 percentage points, p -value < 0.001). Interestingly, the increase in the click rate of product 1 when product 2 is introduced is not only due to some consumers switching from not clicking on anything to clicking on both products but also due to more consumers clicking on *only* product 1. Specifically, 10.75% of consumers in Group 2 click on only product 1, higher than the 10.24% of consumers in Group 1 who click on product 1 (p -value < 0.001). Overall, our data from Groups 1 and 2 suggest that including a new (potentially lower-ranked) product in the choice set could induce more search of the original (potentially higher-ranked) product—and, correspondingly, more purchase of that product—but does not lead to a higher purchase likelihood for that product *conditional* on search.

We also compare the click and purchase rates for products 1 and 2 when a third product is added, but we do not observe evidence of the decoy effect. Specifically, both the click and the purchase rates for product 1 are lower in Group 3 (14.17 percentage points and 4.07 percentage points, respectively) than in Group 2 (16.72 percentage points and 4.74 percentage points, respectively; both p -values < 0.001). Click and purchase rates for product 2 are also lower in Group 3 (7.77 percentage points and 0.36 percentage points, respectively) than in Group 2 (9.98 percentage points and 0.48 percentage points, respectively; both p -values < 0.001). We observe similar patterns when comparing Group 4 to Group 3: adding a fourth product reduces the percentage of consumers clicking or purchasing each of products 1, 2, and 3. Thus, any evidence of the decoy effect that we observe when comparing Groups 1 and 2 does not carry over (or may be dominated by other effects) when the recommendation set size is greater than two.

5.3. Additional Mechanisms

In this section we discuss two additional mechanisms that could potentially lead to reduced search and purchase probabilities from consumers as recommendation set size increases.

Aggressive Sales Effect. One possible reason for the choice overload effect to arise in our setting is that consumers might perceive the recommendation with more products as an overly aggressive sales tactic from the merchant. If this were the case, consumers might switch to purchasing from other merchants on the platform, resulting in demand spillover effects. To investigate whether such spillover effects exist, we obtained data tracking purchases made on Alibaba’s platform by consumers

in our sample both on the day and within 14 days of when they received the recommendation messages. For each consumer i that received a recommendation message from merchant j on day t , we examine two outcome variables: (1) whether she made any purchase from any merchant other than merchant j on the platform on day t (denoted as $PurchaseOtherSeller_{ijt}$), and (2) whether she made any purchase *in the same category as her inquired product* from any merchant other than merchant j on the platform on day t (denoted as $PurchaseOtherSellerSameCate_{ijt}$). We also examine similar outcome variables with time frame extended to within 14 days from day t (denoted as $PurchaseOtherSeller14D_{ijt}$, $PurchaseOtherSellerSameCate14D_{ijt}$).

We conduct regression analyses on these outcome variables using Specification (1). Table A4 in Appendix A presents the regression results. Overall, we find no statistically significant evidence of the demand spillover effect: the likelihood of purchasing from other merchants (either in the same category as the inquired product or across all categories) does not increase when the number of recommended products increases from 2 to 3 and from 3 to 4. This holds regardless of whether we examine the day of or within 14 days of receiving recommendation messages. These results provide side evidence that the mechanism of negative perceptions about merchants pushing a larger recommendation set may not play a major role in driving consumer responses to recommendation set size.

Information Overload in the First Stage. Another possible mechanism that could influence consumers’ search decision as the recommendation set size increases is information overload in the search stage—i.e., the difficulty of deciding which product to *click* on due to cognitive overload. In our field experiment, the effort cost of viewing recommendations in the message was presumably small since the platform ensured that consumers can see the full list of recommended products on one screen, without the need to scroll up and down, regardless of the number of recommended products. Also, the recommendation messages contained relatively limited information about products (i.e., product name, price and picture), which should restrict the role of information overload in the first stage based on the existing literature (Gourville and Soman 2005). However, given that no previous research has empirically examined information overload in the recommender system setting, the presence and extent of information overload at different stages of the recommendation process remains an open question.

6. Heterogeneous Treatment Effects

In this section, we report heterogeneous treatment effects in our data to derive insights into circumstances where consumer engagement is more or less likely to decrease with recommendation set size. Since we observe the choice overload effect only when the number of recommended products exceeds two in our experiment, we mainly investigate the heterogeneous treatment effects of recommending three (as opposed to two) products. Furthermore, we focus on consumers’ search behavior for this analysis, given the importance of click behavior in online recommender systems, the key differentiating role it plays in our hypothesis development, and the direct impact that recommendation set size

has on search behavior (in contrast to purchase behavior, which might be complicated by many other factors in consumers' search process, such as specific product information that consumers acquire in their search).

6.1. Price of the Inquired Product and Time of Inquiry

We first consider two key features of a recommendation set that may influence the effect of set size: the price level of the recommended products and the time of recommendation. Since a recommendation message includes similar products sold by the same merchant, products in a given message tend to fall into the same price range. Thus, we use the price of the inquired product as a proxy for the price level of the options in each recommendation set. For the timing, we focus on whether recommendation messages were sent on a weekday or a weekend. We use the following OLS regression specification to estimate the heterogeneous treatment effects on search probability across the two aforementioned moderators:

$$SearchIncidence_{ijt} = \gamma_0 + \gamma_1 Group_i^3 + \gamma_2 Moderator_{ijt} + \gamma_3 Group_i^3 \times Moderator_{ijt} + X_j + D_t + \epsilon_{ijt}. \quad (2)$$

In Specification (2), $Moderator_{ijt} \in \{\log(Price_{ijt}), Weekday_{ijt}\}$, where $Price_{ijt}$ is the price of the inquired product in the recommendation message sent to consumer i by merchant j on day t , and $Weekday_{ijt}$ is a binary variable that equals 1 if the day t on which merchant j sent the message to consumer i was on a weekday and 0 otherwise. The moderator $\log(Price)$ is standardized to have zero mean and unit deviation before entering the regression. The other variables are the same as those defined in Specification (1). Given our focus on the marginal effect of recommending three (versus two) products, we use the data from Groups 2 and 3 to estimate the model, with Group 2 as the baseline group.

Table A2 in Appendix A reports the estimation results of Specification (2) in relation to the two moderators. First, the marginally significant negative interaction between $\log(Price)$ and the Group 3 indicator suggests that the higher the price level of a recommendation set, the more the search probability is reduced when the number of recommended products increases from two to three. Note that higher prices may be associated with higher anticipated regret (Simonson 1992), in which case the observed pattern is consistent with the predictions of the behavioral model in Appendix C.2 (i.e., Corollary 1). Intuitively, when consumers anticipate greater regret from making an inferior purchase, they may feel more compelled to engage in an extensive search of their options in the recommendation set, and hence are more likely to choose to forgo or defer the search when presented with a larger set. Alternatively, higher prices may be associated with higher search costs (as consumers may pay more attention to product information before making a purchase decision about expensive products). This would again deter consumers from starting a search when presented with a larger recommendation set.

Second, the marginally significant negative interaction between the weekday indicator and the Group 3 indicator suggests that adding a third product reduces the search probability to a *greater* extent if consumers receive the recommendation message on weekdays compared to weekends. This result could be due to consumers having less leisure time—and hence higher search costs—on weekdays than on weekends. That is, when search cost is higher, consumers may find it more difficult to conduct an extensive search and thus become more likely to be deterred as recommendation-set size increases.

6.2. Product Category

Another factor that may drive variances in consumer behavior is product category. Since the inquired product and other recommended products mostly come from the same category, we use the category of the inquired product to represent the category of the options in a recommendation message. In particular, we consider the six largest product categories in our sample: clothing and accessories, home accessories, beauty, baby and children, electronics, and food and drinks. These categories account for more than 75% of our observations, and they are also major product categories on Alibaba’s platform. We find that the size of reduction in search probability caused by adding a third recommended product ranges from 1.19 percentage points to 2.60 percentage points across these six categories (see in Panel A of Table A3 in Appendix A).¹⁸

To evaluate whether the effect of adding a third product on search probability statistically significantly differs across the six categories, we compare two regressions using observations from the six categories in Groups 2 and 3. The first regression predicts search probability as a function of an indicator for Group 3 (vs. Group 2) and indicators for product categories. The other regression additionally includes interactions between product category indicators and the indicator for Group 3 (vs. Group 2). An F -test comparing the two regressions suggests that we can reject the null hypothesis that the effects of adding a third product on search probability are the same across the six major product categories (p-value = 0.0066; see Table A3 in Appendix A for more details).

Altogether, our analysis suggests there exists heterogeneity in the choice overload effect on search behavior across product categories. Consequently, we include product category as a feature in our analysis in Section 6.3, where we use causal forest method to more systematically assess when the choice overload effect is most pronounced.

6.3. Estimation of Heterogeneous Treatment Effects Using Causal Forest Method

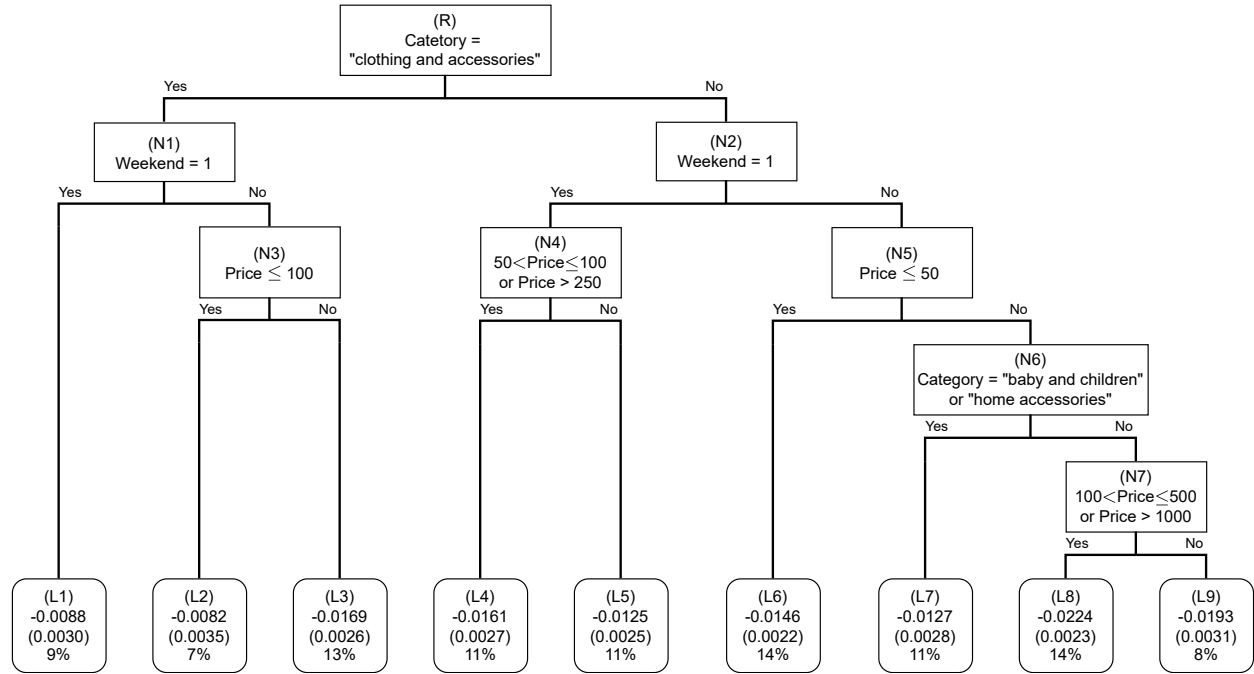
The causal forest method is an emerging technique for analyzing heterogeneity in causal effects, and has been applied in a number of recent empirical studies in economics and operations management

¹⁸ In each of these categories, we also confirm that consumers’ search probability peaks when two products are recommended, and decreases as the recommendation-set size increases further.

(Luo et al. 2019, Guo et al. 2021, Patabhiramaiah et al. 2022, Cui and Davis 2022). It is a semi-parametric approach that extends the classical random forest method by Breiman (2001) and could provide consistent estimates and valid confidence intervals of treatment effects at the individual level (Wager and Athey 2018, Athey et al. 2019) when the data generation process is unconfounded. Given its data-driven, semi-parametric nature, this method enables researchers to discover treatment-effect heterogeneity without predetermining the subgroups for heterogeneity analysis. Because our treatment is randomly assigned and the treatment effect could be heterogeneous across features, the causal forest method is an ideal candidate for us to identify the combinations of key features (product price, inquiry timing, and product category) that may induce a bigger reduction in search probability as the recommendation set grows. The results could shed some light on when and where our main insights from this paper are more important.

We next briefly describe the causal forest method and refer readers to Athey et al. (2019) for more technical details. Consider a dataset of n observations, where each observation $i = 1, \dots, n$ consists of a vector of covariates X_i (which in our case refer to features of recommendation messages), an outcome variable Y_i , and a binary treatment indicator $W_i \in \{0, 1\}$. Let $Y_i(1)$ and $Y_i(0)$ be the potential outcome of unit i when it is assigned to the treatment group ($W_i = 1$) and the control group ($W_i = 0$), respectively. The causal forest method develops an estimator for the conditional average treatment effect, which is defined by the function $\tau(x) = \mathbb{E}[Y_i(1) - Y_i(0) | X_i = x]$. This method consists of two steps. The first step is to construct a large number of causal trees, which establishes for each x a set of neighbors with different weights. Each causal tree is grown using a random subsample of the dataset through recursive partitioning. The splits of the trees are determined in a way to maximize the heterogeneity of the treatment-effect estimates of the child nodes. To reduce the bias of the estimator, the trees must satisfy the “honesty” condition, i.e., no training example can be used for both growing a tree *and* estimating the treatment effects (Wager and Athey 2018). To achieve honesty, the subsample for growing a tree is randomly split into two even and nonoverlapping halves, with one half used to grow the tree and the other half to determine the neighbors of x and their weights for treatment-effect estimation. Then, given x , the second step is to solve a generalized method of moment (GMM) model, using these weighted neighbors of x to obtain $\hat{\tau}(x)$ as the estimate of $\tau(x)$. Athey et al. (2019) prove that the causal-forest estimate $\hat{\tau}(x)$ is consistent and asymptotically Gaussian with closed-form asymptotic variance.

We focus on the heterogeneous effect of recommending three (vs. two) products on consumers’ search behavior. Therefore, our outcome variable is $Y_i = \text{SearchIncidence}_{ijt}$, as defined in Section 5, and the treatment indicator is $W_i = 1$ if consumer i was assigned to Group 3 (i.e., being recommended three products) and $W_i = 0$ if assigned to Group 2 (i.e., being recommended two products). Our covariates X_i contain three variables: (1) the price of the inquired product, Price_{ijt} , as defined in Section 6.1; (2) whether the recommendation message was sent on a weekend, Weekend_{ijt} , as



Note: The top row of each node in the tree is the index of the node. For each leaf, the second, third, and fourth rows report the estimated conditional average treatment effect, standard error, and proportion of observations in the dataset, respectively, of the corresponding subgroup of observations.

Figure 2 Heterogeneous Effect of Adding a Third Recommended Product on Search Probability Estimated by Causal Forest

defined in Section 6.1; and (3) the category of the inquired product referenced in the recommendation message sent to consumer i by merchant j on day t , $Category_{ijt}$.

We conduct our analysis in two steps. First, we train a causal forest using all the observations in Groups 2 and 3. We grow 500 trees in the forest. The hyperparameters of constructing the trees (such as the fraction of data used to build each tree, the number of variables tried for each split, and minimum node size) are determined via cross-validation. Then for each observation i with covariates X_i , we use the trained forest to obtain the out-of-bag prediction (i.e., prediction made by using only the trees that do not use observation i for growing them) for the conditional average treatment effect $\hat{\tau}^{(-i)}(X_i)$. Second, we construct a regression tree with $\hat{\tau}^{(-i)}(X_i)$ as the outcome variable and X_i as predictors.¹⁹ The tree splits are determined by maximizing the between-group variation of the two nodes after the splits. We then estimate the conditional average treatment effect and its standard error for each leaf using the doubly robust estimator as in Equation (8) of Athey and Wager (2019).

The results of our causal-forest analysis, depicted in Figure 2, suggest substantial heterogeneity in the conditional average treatment effect of adding a third product on search probability across recommendation messages with different features: the magnitude of reduction in search probability ranges

¹⁹ For ease of demonstrating and interpreting the results, in the second step, we construct a categorical variable that includes seven price levels: $(0, 10]$, $(10, 50]$, $(50, 100]$, $(100, 250]$, $(250, 500]$, $(500, 1000]$, and $(1000, +\infty)$ (units are all in CNY). We use this categorical variable instead of the original price as the predictor for constructing the regression tree.

from 0.82 percentage points to 2.24 percentage points across the nine subgroups of features identified in our analysis. For example, consumers who inquire about products in the “clothing and accessories” category tend to exhibit the smallest choice overload effect on search probability—particularly on weekends (leaf L1 in Figure 2, which has the second-smallest treatment effect overall) or when the inquired product is priced below 100 CNY (leaf L2 in Figure 2, which has the smallest treatment effect overall). This observation could be because when shopping for clothing and accessories (versus other categories of products), consumers enjoy the search process more (Arnold and Reynolds 2003, Babin et al. 1994) and hence are less likely to defer the search when the recommendation set size increases. The enjoyment of search may play a more dominant role (and thus the treatment effect of adding a third product to the recommendation set is smaller) on weekends—when consumers may have more leisure time and hence lower search costs—and for products that cost less than 100 CNY—when consumers may be less concerned about making the “right” decision (see also the discussion in Section 6.1 on the effects of time and price). Among the remaining categories, the treatment effect is generally more pronounced on weekdays, but it still varies across different product categories and for different price ranges. The largest treatment effects are observed when the inquiry is made on a weekday about a product costing more than 50 CNY that belongs to the “beauty,” “electronics,” or “food and drink” categories (leaves L8 and L9 in Figure 2, where the product does *not* belong to the categories listed in node R or N6).

Overall, our results suggest that the treatment effect of adding a third product to the recommendation set can vary depending on the combination of the timing of recommendation as well as the price and category of recommended products. By identifying the specific feature combinations that yield the highest or lowest treatment effects, our analysis in this section illustrates the value of applying data-mining approaches to studying heterogeneous treatment effects with multiple features. Similar causal-forest approaches could help researchers and practitioners pinpoint products that may benefit the most from keeping the recommendation set small (or from other interventions designed to counter the choice overload effect).

7. Discussion and Conclusion

In this paper, we provide, to the best of our knowledge, the first empirical evidence of the choice overload effect in the context of online recommender systems. Via a large-scale field experiment, we show that including more products in a recommendation set can lead to a lower purchase probability from consumers. Moreover, we show that this reduction in purchase probability is largely driven by a reduction in consumers’ likelihood of searching the recommendation set (by clicking on at least one of the recommended products). We discuss the possible behavioral mechanisms driving these effects, and we also explore heterogeneous treatment effects in our data to identify the conditions under which offering a larger recommendation set is most likely to discourage consumer search.

7.1. Managerial Implications

Our research yields several managerial implications for retailers and retailing platforms. First, we demonstrate the existence of the choice overload effect in a common online retail setting—the recommender system of a retargeting channel—and we show that consumers’ search and purchase probabilities first increase then *decrease* as the recommendation set size increases. Therefore, though it may be tempting for online retailers to recommend more options to consumers—especially given the low display costs online—they should be cautious about doing so. Careful planning and testing is needed to identify the recommendation set size that will lead to the largest profit.

Second, our results provide suggestive evidence for the heterogeneity in how a larger recommendation set influences consumers’ likelihood of starting a search, suggesting that the choice overload effect on search probability may be more pronounced for more expensive products and on weekdays and may differ across product categories. These observations suggest that, for retailers to encourage consumer search, they may consider focusing not only on how to personalize product offerings for a given recommendation set size but also on dynamically personalizing the set size—for example, by offering different sizes for different product categories, price ranges, and times.

Third, our results yield important implications for firms that wish to mitigate the choice overload effect in their recommender systems. In particular, prior laboratory studies have found that, to attenuate the choice overload effect, retailers can change the information display format or the amount of information they give to consumers (Gourville and Soman 2005, Townsend and Kahn 2014). However, these prior studies have typically exposed subjects to all available information about products in the choice set. As such, their findings cannot be directly applied to recommendation settings where detailed product information is accessible to consumers only *after* they decide to search (click on) the products. In fact, our experiment shows that a larger recommendation set can reduce the likelihood that consumers will start a search in the first place, which in turn substantially reduces their purchase likelihood. This finding suggests that altering the presentation of information on product pages, which are visible to consumers only after searching, might not be sufficient for curbing the choice overload effect. Rather, firms should focus on improving the design of the recommender system interface that first presents consumers with a recommendation set.

7.2. Robustness

To check the robustness of our main empirical findings, we conduct several additional analyses, which we briefly describe below.

First, different merchants on the platform may have different scopes of offerings, which would affect the similarity among the recommended products. To investigate whether this factor influences the main effect we observe, we classify the recommendation messages into two types based on whether or not the inquired product and the second recommended product (i.e., the product with the highest

similarity to the inquired product, as determined by the algorithm) are in the same sub-category.²⁰ Note that since the second recommended product is most similar to the inquired product, if the second recommended product is not in the same sub-category as the inquired product, the third and fourth recommended products in Groups 3 and 4 would also be unlikely to be in the same sub-category as the inquired product. We observe that as recommendation set size increases from two to three then to four, click and purchase probabilities consistently decrease, regardless of whether or not the inquired product and the second recommended product are in the same sub-category. These results suggest that in our dataset, the choice overload effect is not significantly impacted by the similarity among recommended products.²¹ Please see Appendix B.1 for details of the analysis.

Second, consumers' responses to recommendation messages may depend on whether or not they need to start a new shopping session, which could be correlated with the *promptness* with which consumers engage with a recommendation message. To investigate this question, we construct two outcome variables to capture whether each consumer clicked on any product within 4.8 minutes of receiving a recommendation message and whether each consumer clicked on any product after 4.8 minutes (where 4.8 minutes is the median value of how long it took consumers to click after a recommendation message was sent during our experiment). We show that the inverted-U patterns exist for both outcome measures of click behavior, and that consumers' purchasing behavior conditional on clicking was similar regardless of whether consumers clicked within 4.8 minutes or after 4.8 minutes. These observations provide suggestive evidence that in our empirical setting, the documented choice overload effects may be robust to how quickly consumers engage with the messages. See Appendix B.2 for details.

Finally, we explore the extent to which consumers in our analysis represent the general consumer population on Alibaba's retailing platform in terms of purchasing behaviors (see Appendix B.3 for details). To summarize, using consumer purchasing data in the month prior to our experiment, we find that the distribution modes and skewness of order frequency and order amount are largely comparable between consumers in our experiment and general consumers on the platform. We do notice that a larger proportion of consumers in our experiment have high order frequencies and/or high order amounts, compared to the general consumer population. This suggests that while our

²⁰ Recommended products are mostly in the same general category, but may not be in the same sub-category. The platform has multiple levels of sub-categories. For this analysis, we use the finest level, i.e., the leaf category. Products in the same leaf category not only have the same product function but also share some other product characteristics, such as targeted consumer group, material, style, specialized features, etc. Some examples of the leaf categories include "men's cotton pants", "women's lace-up boots", "wireless sports earbuds."

²¹ Since all recommendation messages in Groups 2, 3, and 4 contain the inquired product and the second recommended product, our classification based on the sub-category alignment between the second recommended product and the inquired product guarantees comparability across these three experimental groups. If we change the classification criterion to capture whether all other recommended products are in the same sub-category as the inquired product, we also find that the effect of recommendation set size does not differ statistically significantly when all recommended products are in the same sub-category versus when they are not.

experiment may have moderately oversampled the more active purchasers on the platform, the consumers in our experiment by and large represent the general consumer population in terms of diversity in purchasing behavior. These findings, combined with the fact that a significant proportion (35%) of consumers on the platform inquire with a merchant before purchasing from the merchant, indicate that our empirical findings can be quite relevant for the platform.

7.3. Directions for Future Research

Our work has several limitations, which, we believe, open up future research directions. To begin with, as the first step toward empirically investigating the choice overload effect in an online retail setting, we conducted our field experiment through the product-recommendation service via a customer-service messenger on an online retail platform. Our experiment took place in the platform’s retargeting campaign, which specifically targeted consumers who indicated interest in a product (by making an inquiry) but did not make a purchase within a certain time period. As we discussed in Section 2, this is a context where personalized recommendation messages are likely to be especially effective, and so any moderator of that effectiveness could be particularly relevant for practice. However, it would also be important to study whether our findings generalize to other types of recommender system contexts, such as when platforms target a broader consumer population who may not have interacted with a specific product (i.e., those earlier in the purchasing funnel), or if the recommendations are sent *while* consumers browse products on the websites (e.g., via display banners on product pages). Moreover, our empirical setting involves recommending products from the same merchant that the consumer made an inquiry with, which is common among single-brand retailers or within a merchant’s private domain (e.g., a merchant’s store page or its product pages) on a platform. A valuable future direction is to explore whether our empirical findings remain consistent when recommended products come from different merchants.

Beyond extending our research to other recommender system contexts, there are several additional interesting research directions. First, we focus on consumers’ purchase behavior and only consider purchases made within the recommendation channel, which reflects our field partner’s interest in conducting this research. However, it would be valuable for future research to inspect how overall purchases consumers make across multiple channels as well as post-purchase decisions (e.g., return rates) or other long-term measures (e.g., customer retention) vary with the recommendation set size. Second, future research could evaluate whether the text content of consumers’ inquiries and merchants’ recommendation messages—which we do not have access to in our setting—can act as moderating or mediating factors of the main effects we observe. Third, as discussed in Section 5.3, there may be alternative mechanisms contributing to the choice overload effect in recommender systems, and thus future research could further explore the extent to which these mechanisms are present in different settings. In addition, given that our experiment is conducted in a standard

retailing app interface, it would be valuable to empirically test how interventions such as altering the interface design of recommender systems shape the relationship between recommendation set size, search, and purchase behavior. Finally, future research could leverage our findings on consumer behavior to study efficient estimation and optimization algorithms that jointly determine the optimal recommendation set size as well as the products to include.

References

- Abendroth LJ, Diehl K (2006) Now or never: Effects of limited purchase opportunities on patterns of regret over time. *Journal of Consumer Research* 33(3):342–351.
- Arnold MJ, Reynolds KE (2003) Hedonic shopping motivations. *Journal of Retailing* 79(2):77–95.
- Athey S, Tibshirani J, Wager S (2019) Generalized random forests. *The Annals of Statistics* 47(2):1148–1178.
- Athey S, Wager S (2019) Estimating treatment effects with causal forests: An application. *Observational Studies* 5(2):37–51.
- Babin BJ, Darden WR, Griffin M (1994) Work and/or fun: Measuring hedonic and utilitarian shopping value. *Journal of Consumer Research* 20(4):644–656.
- Bell DE (1982) Regret in decision making under uncertainty. *Operations Research* 30(5):961–981.
- Bleier A, Eisenbeiss M (2015) Personalized online advertising effectiveness: The interplay of what, when, and where. *Marketing Science* 34(5):669–688.
- Bobadilla J, Ortega F, Hernando A, Gutiérrez A (2013) Recommender systems survey. *Knowledge-based Systems* 46:109–132.
- Bollen D, Knijnenburg BP, Willemsen MC, Graus M (2010) Understanding choice overload in recommender systems. *Proceedings of the 4th ACM Conference on Recommender Systems*, 63–70.
- Breiman L (2001) Random forests. *Machine Learning* 45(1):5–32.
- Buell RW, Kalkanci B (2021) How transparency into internal and external responsibility initiatives influences consumer choice. *Management Science* 67(2):932–950.
- Cachon GP, Gallino S, Olivares M (2019) Does adding inventory increase sales? Evidence of a scarcity effect in US automobile dealerships. *Management Science* 65(4):1469–1485.
- Calvo E, Cui R, Wagner L (2023) Disclosing product availability in online retail. *Manufacturing & Service Operations Management* 25(2):427–447.
- Chen Y, Yao S (2017) Sequential search with refinement: Model and application with click-stream data. *Management Science* 63(12):4345–4365.
- Chernev A, Böckenholt U, Goodman J (2015) Choice overload: A conceptual review and meta-analysis. *Journal of Consumer Psychology* 25(2):333–358.
- Cui R, Zhang DJ, Bassamboo A (2019) Learning from inventory availability information: Evidence from field experiments on Amazon. *Management Science* 65(3):1216–1235.
- Cui Y, Davis AM (2022) Tax-induced inequalities in the sharing economy. *Management Science* 68(10):7202–7220.
- De los Santos B, Hortaçsu A, Wildenbeest MR (2012) Testing models of consumer search using data on web browsing and purchasing behavior. *American Economic Review* 102(6):2955–2980.
- De los Santos B, Koulayev S (2017) Optimizing click-through in online rankings with endogenous search refinement. *Marketing Science* 36(4):542–564.
- Dean M, Ravindran D, Stoye J (2022) A better test of choice overload.
- Diehl K, Poynor C (2010) Great expectations?! Assortment size, expectations, and satisfaction. *Journal of Marketing Research* 47(2):312–322.
- Feldman J, Zhang DJ, Liu X, Zhang N (2022) Customer choice models vs. machine learning: Finding optimal product displays on Alibaba. *Operations Research* 70(1):309–328.
- Ferreira KJ, Lee BHA, Simchi-Levi D (2016) Analytics for an online retailer: Demand forecasting and price optimization. *Manufacturing & Service Operations Management* 18(1):69–88.
- Fisher M, Gallino S, Li J (2018) Competition-based dynamic pricing in online retailing: A methodology validated with field experiments. *Management Science* 64(6):2496–2514.
- Gallego G, Li A, Truong VA, Wang X (2020) Approximation algorithms for product framing and pricing. *Operations Research* 68(1):134–160.
- Gallino S, Karacaoglu N, Moreno A (2023) Need for speed: The impact of in-process delays on customer behavior in online retail. *Operations Research* 71(3):876–894.
- Gallino S, Moreno A (2018) The value of fit information in online retail: Evidence from a randomized field experiment. *Manufacturing & Service Operations Management* 20(4):767–787.
- Ghose A, Ipeirotis PG, Li B (2012) Designing ranking systems for hotels on travel search engines by mining user-generated and crowdsourced content. *Marketing Science* 31(3):493–520.

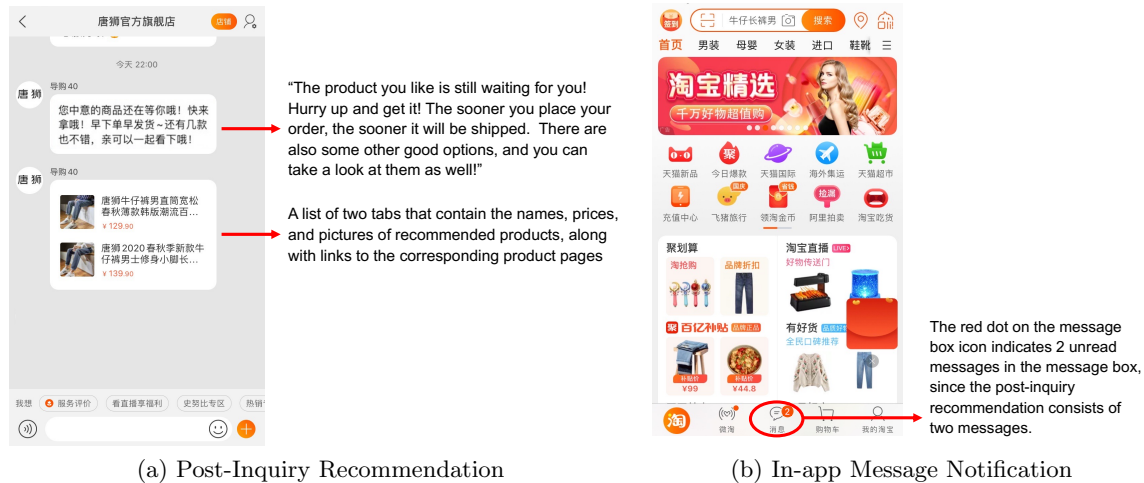
- Gourville JT, Soman D (2005) Overchoice and assortment type: When and why variety backfires. *Marketing Science* 24(3):382–395.
- Greifeneder R, Scheibehenne B, Kleber N (2010) Less may be more when choosing is difficult: Choice complexity and too much choice. *Acta Psychologica* 133(1):45–50.
- Guo T, Sriram S, Manchanda P (2021) The effect of information disclosure on industry payments to physicians. *Journal of Marketing Research* 58(1):115–140.
- Hadar L, Sood S (2014) When knowledge is demotivating: Subjective knowledge and choice overload. *Psychological Science* 25(9):1739–1747.
- Hauser JR (2014) Consideration-set heuristics. *Journal of Business Research* 67(8):1688–1699.
- Hoban PR, Bucklin RE (2015) Effects of Internet display advertising in the purchase funnel: Model-based insights from a randomized field experiment. *Journal of Marketing Research* 52(3):375–393.
- Honka E (2014) Quantifying search and switching costs in the US auto insurance industry. *The RAND Journal of Economics* 45(4):847–884.
- Honka E, Chintagunta P (2017) Simultaneous or sequential? Search strategies in the US auto insurance industry. *Marketing Science* 36(1):21–42.
- Hosanagar K, Fleder D, Lee D, Buja A (2014) Will the global village fracture into tribes? Recommender systems and their effects on consumer fragmentation. *Management Science* 60(4):805–823.
- Hu K, Wang YI (2022) More or less: How information richness affects our decision consistency.
- Inbar Y, Botti S, Hanks K (2011) Decision speed and choice regret: When haste feels like waste. *Journal of Experimental Social Psychology* 47(3):533–540.
- Irons B, Hepburn C (2007) Regret theory and the tyranny of choice. *Economic Record* 83(261):191–203.
- Iyengar SS, Lepper MR (2000) When choice is demotivating: Can one desire too much of a good thing? *Journal of Personality and Social Psychology* 79(6):995–1006.
- Jacoby J, Speller DE, Berning CK (1974) Brand choice behavior as a function of information load: Replication and extension. *Journal of Consumer Research* 1(1):33–42.
- Jiang ZZ, Kong G, Zhang Y (2021) Making the most of your regret: Workers' relocation decisions in on-demand platforms. *Manufacturing & Service Operations Management* 23(3):695–713.
- Johnson GA, Lewis RA, Nubbemeyer EI (2017) Ghost ads: Improving the economics of measuring online ad effectiveness. *Journal of Marketing Research* 54(6):867–884.
- Khern-am nuai W, So H, Cohen MC, Adulyasak Y (2024) Selecting cover images for restaurant reviews: AI vs. wisdom of the crowd. *Manufacturing & Service Operations Management* 26(1):330–349.
- Kim JB, Albuquerque P, Bronnenberg BJ (2010) Online demand under limited consumer search. *Marketing Science* 29(6):1001–1023.
- Kuksov D, Villas-Boas JM (2010) When more alternatives lead to less choice. *Marketing Science* 29(3):507–524.
- Lambrecht A, Tucker C (2013) When does retargeting work? Information specificity in online advertising. *Journal of Marketing Research* 50(5):561–576.
- Lee BK, Lee WN (2004) The effect of information overload on consumer choice quality in an on-line environment. *Psychology & Marketing* 21(3):159–183.
- Levav J, Heitmann M, Herrmann A, Iyengar SS (2010) Order in product customization decisions: Evidence from field experiments. *Journal of Political Economy* 118(2):274–299.
- Li J, Luo X, Lu X, Moriguchi T (2021) The double-edged effects of e-commerce cart retargeting: Does retargeting too early backfire? *Journal of Marketing* 85(4):123–140.
- Li S, Lu LX, Lu SF, Huang S (2023) Estimating the stockout-based demand spillover effect in a fashion retail setting. *Manufacturing & Service Operations Management* 25(2):468–488.
- Loomes G, Sugden R (1982) Regret theory: An alternative theory of rational choice under uncertainty. *The Economic Journal* 92(368):805–824.
- Lu B, Xu Y, Dai H, Zhou W (2022) When is personalized price discount more efficient? A quasi-experiment study.
- Lü L, Medo M, Yeung CH, Zhang YC, Zhang ZK, Zhou T (2012) Recommender systems. *Physics Reports* 519(1):1–49.
- Lu Y, Musalem A, Olivares M, Schilkut A (2013) Measuring the effect of queues on customer purchases. *Management Science* 59(8):1743–1763.
- Luce RD (1959) *Individual choice behavior: A theoretical analysis* (New York, Wiley).
- Luo X, Lu X, Li J (2019) When and how to leverage e-commerce cart targeting: The relative and moderated effects of scarcity and price incentives with a two-stage field experiment and causal forest optimization. *Information Systems Research* 30(4):1203–1227.
- Moon K, Bimpikis K, Mendelson H (2018) Randomized markdowns and online monitoring. *Management Science* 64(3):1271–1290.
- Musalem A, Olivares M, Bradlow ET, Terwiesch C, Corsten D (2010) Structural estimation of the effect of out-of-stocks. *Management Science* 56(7):1180–1197.

- Ngwe D, Ferreira KJ, Teixeira T (2019) The impact of increasing search frictions on online shopping behavior: Evidence from a field experiment. *Journal of Marketing Research* 56(6):944–959.
- Pattabhiramaiah A, Overby E, Xu L (2022) Spillovers from online engagement: How a newspaper subscriber’s activation of digital paywall access affects her retention and subscription revenue. *Management Science* 68(5):3528–3548.
- Pires T (2016) Costly search and consideration sets in storable goods markets. *Quantitative Marketing and Economics* 14(3):157–193.
- Pocheptsova A, Amir O, Dhar R, Baumeister RF (2009) Deciding without resources: Resource depletion and choice in context. *Journal of Marketing Research* 46(3):344–355.
- Reutskaja E, Hogarth RM (2009) Satisfaction in choice as a function of the number of alternatives: When “goods satiate”. *Psychology & Marketing* 26(3):197–203.
- Rieskamp J, Busemeyer JR, Mellers BA (2006) Extending the bounds of rationality: Evidence and theories of preferential choice. *Journal of Economic Literature* 44(3):631–661.
- Rios I, Saban D, Zheng F (2022) Improving match rates in dating markets through assortment optimization. *Manufacturing & Service Operations Management* .
- Sagi A, Friedland N (2007) The cost of richness: The effect of the size and diversity of decision sets on post-decision regret. *Journal of Personality and Social Psychology* 93(4):515.
- Sahni NS, Narayanan S, Kalyanam K (2019) An experimental investigation of the effects of retargeted advertising: The role of frequency and timing. *Journal of Marketing Research* 56(3):401–418.
- Sarver T (2008) Anticipating regret: Why fewer options may be better. *Econometrica* 76(2):263–305.
- Scheibehenne B, Greifeneder R, Todd PM (2010) Can there ever be too many options? A meta-analytic review of choice overload. *Journal of Consumer Research* 37(3):409–425.
- Schrage M (2017) Great digital companies build great recommendation engines. *Harvard Business Review* .
- Schwartz B, Ward A, Monterosso J, Lyubomirsky S, White K, Lehman DR (2002) Maximizing versus satisficing: Happiness is a matter of choice. *Journal of Personality and Social Psychology* 83(5):1178.
- Simonsohn U, Nelson LD, Simmons JP (2014) p-curve and effect size: Correcting for publication bias using only significant results. *Perspectives on Psychological Science* 9(6):666–681.
- Simonson I (1992) The influence of anticipating regret and responsibility on purchase decisions. *Journal of Consumer Research* 19(1):105–118.
- Stamatopoulos I, Bassamboo A, Moreno A (2021) The effects of menu costs on retail performance: Evidence from adoption of the electronic shelf label technology. *Management Science* 67(1):242–256.
- Stigler GJ (1961) The economics of information. *Journal of Political Economy* 69(3):213–225.
- Talluri K, Van Ryzin G (2004) Revenue management under a general discrete choice model of consumer behavior. *Management Science* 50(1):15–33.
- Timmermans D (1993) The impact of task complexity on information use in multi-attribute decision making. *Journal of Behavioral Decision Making* 6(2):95–111.
- Townsend C, Kahn BE (2014) The “visual preference heuristic”: The influence of visual versus verbal depiction on assortment processing, perceived variety, and choice overload. *Journal of Consumer Research* 40(5):993–1015.
- Ursu RM (2018) The power of rankings: Quantifying the effect of rankings on online consumer search and purchase decisions. *Marketing Science* 37(4):530–552.
- Wager S, Athey S (2018) Estimation and inference of heterogeneous treatment effects using random forests. *Journal of the American Statistical Association* 113(523):1228–1242.
- Wang R, Sahin O (2018) The impact of consumer search cost on assortment planning and pricing. *Management Science* 64(8):3649–3666.
- Wu C, Cosguner K (2020) Profiting from the decoy effect: A case study of an online diamond retailer. *Marketing Science* 39(5):974–995.
- Zeelenberg M (1999) The use of crying over spilled milk: A note on the rationality and functionality of regret. *Philosophical Psychology* 12(6):325–340.
- Zhang DJ, Dai H, Dong L, Qi F, Zhang N, Liu X, Yang J (2020) The long-term and spillover effects of price promotions on retailing platforms: Evidence from a large randomized experiment on Alibaba. *Management Science* 66(6):2589–2609.
- Zhang S, Yao L, Sun A, Tay Y (2019) Deep learning based recommender system: A survey and new perspectives. *ACM Computing Surveys (CSUR)* 52(1):1–38.

Appendix. The Choice Overload Effect in Online Recommender Systems

Xiaoyang Long, Jiankun Sun, Hengchen Dai, Dennis J. Zhang,
Jianfeng Zhang, Yujie Chen, Haoyuan Hu, Binqiang Zhao

A. Supplementary Figures and Tables



(a) Post-Inquiry Recommendation (b) In-app Message Notification
Figure A1 Examples of a Post-Inquiry Recommendation and an In-App Message Notification

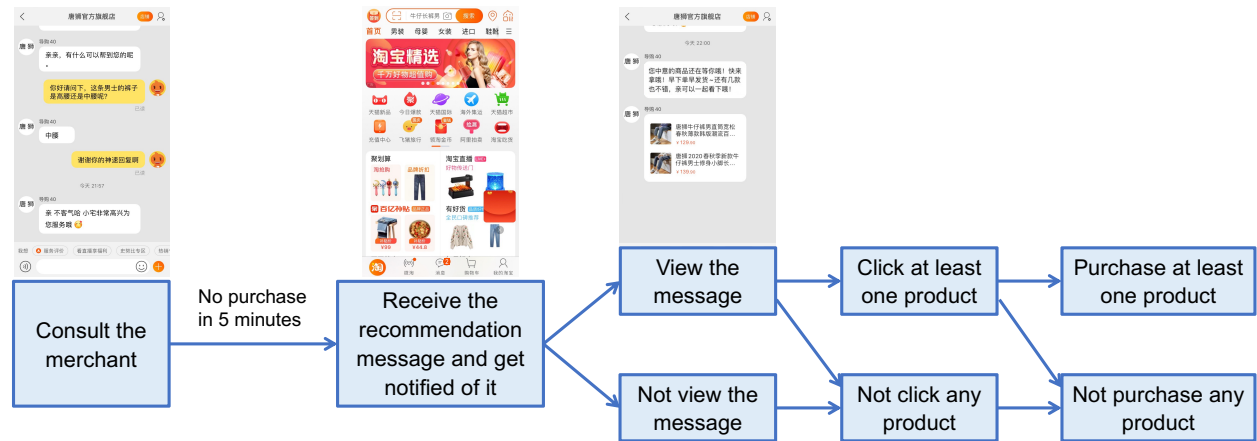


Figure A2 Flow of a Typical Consumer's Experiences and Actions in Response to the Recommendation Message

Table A1 Summary Statistics and Randomization Checks

	Group 1	Group 2	Group 3	Group 4	Minimum P-value Across Six Pairwise Comparisons
Number of Viewed Messages	402,770	405,178	403,733	403,644	
log(Inquired Product Price)	4.87	4.87	4.87	4.87	> 0.69
Percentage of Weekend Messages	31.78%	31.70%	31.75%	31.84%	> 0.17

Notes: This table reports the number of viewed messages, average log-transformed inquired product price in Chinese Yuan, and the percentage of messages sent on weekends for each experimental group. The last column of the table reports the minimum p-value of the six pairwise t-tests on log-transformed inquired product price and the six pairwise two-sample proportion tests on proportions of messages sent on weekends across groups. The p-values of pairwise statistical tests for the differences in the above metrics are all greater than 0.17.

Table A2 Heterogeneous Treatment Effects of Adding a Third Recommended Product, by Price and Recommendation Timing

Outcome Variable	SearchIncidence	
	Moderator = log(Price) (1)	Moderator = Weekday (2)
Group 3	−0.0152**** (0.0009)	−0.0136**** (0.0015)
Moderator	0.0115**** (0.0012)	
Group 3 × Moderator	−0.0016* (0.0009)	−0.0033* (0.0019)
Merchant Fixed Effects	Yes	Yes
Date Fixed Effects	Yes	Yes
Observations	808,911	808,911

Notes: We use only the observations in Group 2 and Group 3 in the regressions. Columns (1) and (2) report the heterogeneous treatment effects of recommending three (vs. two) products, moderated by the logarithm of the inquired product's price and whether the recommendation message was sent on weekdays or not, respectively. The moderator log(Price) is standardized to have zero mean and unit standard deviation before entering the regression. In Column (2), there is no estimate for the moderator *Weekday* due to the inclusion of date fixed effects in the regression. Standard errors are robust and clustered at the merchant level. *p<0.1; **p<0.05; ***p<0.01; ****p<0.001.

Table A3 Treatment Effect of Adding a Third Recommended Product, by Major Product Category

Category	Outcome Variable = SearchIncidence					
	Clothing and Accessories (1)	Home Accessories (2)	Beauty (3)	Baby and Children (4)	Electronics (5)	Food and Drinks (6)
Group 3	−0.0132**** (0.0017)	−0.0133**** (0.0023)	−0.0176**** (0.0027)	−0.0119**** (0.0032)	−0.0260**** (0.0030)	−0.0152**** (0.0032)
Merchant Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Date Fixed Effects	Yes	Yes	Yes	Yes	Yes	Yes
Observations	235,426	117,759	77,522	73,717	58,703	51,021
Proportion of Observations	29.1%	14.6%	9.6%	9.1%	7.3%	6.3%

Notes: This table reports the estimated treatment effect of recommending three (vs. two) products on search probability for each major product category. We use the following specification for estimation: $SearchIncidence_{ijt} = \beta_0 + \beta_1 Group_i^3 + X_j + D_t + \epsilon_{ijt}$, with all variables defined as in Specification (1) in the paper. We only use the observations in Group 2 and Group 3 for the analyses. The last row reports the proportion of total observations in the experiment for each category. Standard errors are robust and clustered at the merchant level. *p<0.1; **p<0.05; ***p<0.01; ****p<0.001.

To examine whether the treatment effects are equal across the six major product categories, we conduct an *F*-test with the null hypothesis that $\theta_3^k = 0$ for $k = 1, 2, 3, 4, 5$ in the following OLS regression specification: $SearchIncidence_{ijt} = \theta_0 + \theta_1 Group_i^3 + \sum_{k=1}^5 \theta_2^k Category_{ijt}^k + \sum_{k=1}^5 \theta_3^k Group_i^3 \times Category_{ijt}^k + X_j + D_t + \epsilon_{ijt}$, where $Category_{ijt}^k$ is a binary variable that equals 1 if the inquired product in the recommendation message sent to consumer *i* by merchant *j* on day *t* is in the category *k* and 0 otherwise, for $k = 1, 2, 3, 4, 5$ (each category is indexed by its column number in the above table). Category 6 (i.e., food and drink) serves as the baseline category in the regression specification. The p-value of the *F*-test is 0.0066 (*F*-statistics = 3.2157), and thus we reject the null hypothesis that $\theta_3^k = 0$ for $k = 1, 2, 3, 4, 5$. In other words, these treatment effects are not statistically equal across the six categories.

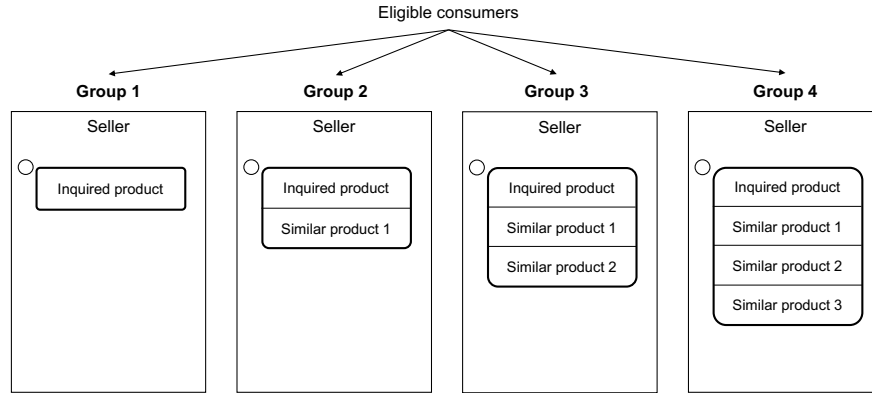


Figure A3 Outline of Recommendation Messages Sent to the Four Experimental Groups

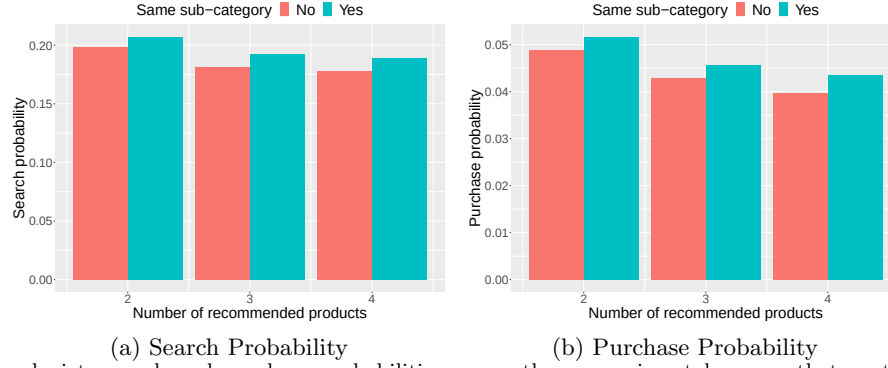
Table A4 Consumer Purchase From Other Merchants

Panel A: Consumer Purchase On the Day of Receiving a Recommendation Message				
Outcome Variable	PurchaseOtherSeller		PurchaseOtherSellerSameCate	
	Baseline: Group 2 (1)	Baseline: Group 3 (2)	Baseline: Group 2 (3)	Baseline: Group 3 (4)
Group 1	-0.0002 (0.0011)	0.0014 (0.0011)	0.0013 (0.0009)	0.0013 (0.0009)
Group 2		0.0016 (0.0011)		-0.0001 (0.0009)
Group 3	-0.0016 (0.0011)		0.0001 (0.0009)	
Group 4	-0.0021* (0.0011)	-0.0005 (0.0011)	-0.0007 (0.0009)	-0.0008 (0.0006)
Panel B: Consumer Purchase Within 14 Days of Receiving a Recommendation Message				
Outcome Variable	PurchaseOtherSeller14D		PurchaseOtherSellerSameCate14D	
	Baseline: Group 2 (1)	Baseline: Group 3 (2)	Baseline: Group 2 (3)	Baseline: Group 3 (4)
Group 1	-0.0001 (0.0006)	0.0004 (0.0006)	0.0008 (0.0011)	0.0010 (0.0011)
Group 2		0.0005 (0.0006)		0.0003 (0.0011)
Group 3	-0.0005 (0.0006)		-0.0003 (0.0011)	
Group 4	-0.0003 (0.0006)	0.0002 (0.0006)	-0.0011 (0.0011)	-0.0009 (0.0011)
Merchant Fixed Effects	Yes	Yes	Yes	Yes
Date Fixed Effects	Yes	Yes	Yes	Yes
Observations	1,615,325	1,615,325	1,615,325	1,615,325

Notes: The table reports results from OLS regressions where standard errors are robust and clustered at the merchant level. Panel A reports the estimated differences between experimental groups in the probabilities of consumers purchasing from other merchants across all categories (columns (1) and (2)) or from the same category as the inquired product (columns (3) and (4)) on the day of receiving a recommendation message. Panel B follows a similar structure, but the time frame extends to within 14 days of receiving a recommendation message. In both panels, columns (1) and (3) have Group 2 as the baseline group, and columns (2) and (4) have Group 3 as the baseline group. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$; **** $p < 0.001$.

B. Additional Data Analysis

In this appendix, we provide details of additional analyses mentioned in the paper.



Notes: This figure depicts search and purchase probabilities across three experimental groups that contain more than one recommended product, broken down by whether the second recommended product is in the same sub-category as the inquired product or not.

Figure A4 Search Probability and Purchase Probability As a Function of Recommendation Set Size, Broken Down by Product Sub-category Alignment

B.1. Similarity of Recommended Products

We leverage whether the inquired product and the second recommended product (i.e., the product with the highest similarity to the inquired product, as determined by the algorithm) are in the same sub-category as a measurement of the similarity of products within each recommendation message. Specifically, we classify recommendation messages in Groups 2, 3, and 4 into two types, depending on whether the second recommended product is in the same sub-category as the inquired product (i.e., recommendations with greater similarity) or not (i.e., recommendations with lesser similarity). Since all recommendation messages in Groups 2, 3, and 4 contain the inquired product and the second recommended product, our classification based on the sub-category alignment between the second recommended product and the inquired product guarantees comparability across these three experimental groups.²² Figure A4 and Table A5 present how the click and purchase probabilities vary with the number of recommended products, separately for when the second recommended product and the inquired product are in the same sub-category versus when they are not. For both types of messages, as recommendation set size increases from two to three and then from three to four, click and purchase probabilities decrease statistically significantly.²³

²² An alternative classification criterion is to look at whether all other recommended products are in the same sub-category as the inquired product. With this criterion, we also find that the effect of recommendation set size on click and purchase probabilities does not differ when all recommended products are in the same sub-category versus when they are not. We do not focus on this approach here because recommendation messages with full alignment (or without full alignment) across recommended products may not be comparable between experimental groups, given that the number of recommended products differs.

²³ We observe that the click and purchase probabilities are generally higher when the second recommended product and the inquired product are in the same sub-category than when they are not. This may not be a causal comparison but is as expected, since recommendations in the same sub-category as the inquired product may align more closely with consumers' preferences compared to recommendations from different sub-categories.

Table A5 Summary Statistics of Search Probability and Purchase Probability as a Function of Recommendation Set Size, Broken Down by Product Sub-category Alignment

	Group 2 (2 products)	Group 3 (3 products)	Group 4 (4 products)	Absolute Difference Between Groups 2 and 3 and P-value	Absolute Difference Between Groups 3 and 4 and P-value
Panel A: Search Probability					
Same Sub-category	20.73%	19.24%	18.94%	−1.49% (< 0.001)	−0.30% (0.003)
Not Same Sub-category	19.80%	18.14%	17.76%	−1.66% (< 0.001)	−0.38% (0.019)
Panel B: Purchase Probability					
Same Sub-category	5.17%	4.57%	4.34%	−0.60% (< 0.001)	−0.23% (< 0.001)
Not Same Sub-category	4.88%	4.29%	3.97%	−0.59% (< 0.001)	−0.32% (< 0.001)

Notes: This table compares search and purchase probabilities across three experimental groups that contain more than one recommended product, broken down by whether the second recommended product is in the same sub-category as the inquired product ("Same Sub-category") or not ("Not Same Sub-category"). We report p-values of two-sample proportion tests for all comparisons in this table.

For further analysis, we run the following linear regression analysis to examine whether the effect of recommendation set size significantly differs when the inquired product and the second recommended product are in the same sub-category versus when they are not:

$$\begin{aligned}
 OutcomeVariable_{ijt} = & \alpha_0 + \sum_{n \in \{2,3,4\}} \alpha_1^n Group_i^n + \alpha_2 SameCategory_i \\
 & + \sum_{n \in \{2,3,4\}} \alpha_3^n Group_i^n \times SameCategory_i + X_j + D_t + \epsilon_{ijt}.
 \end{aligned}$$

where $SameCategory_i$ equals 1 if the inquired product and the second recommended product received by consumer i are in the same sub-category and 0 otherwise. All other variable definitions remain consistent with Specification (1) in the paper. Table A6 shows that the interactions between the indicators of experimental groups and $SameCategory_i$ are not statistically significant across the board. This suggests that there are no statistically significant differences in the treatment effects of recommendation set size on click and purchase behaviors between these two types of messages. These results provide suggestive evidence that in our dataset, the similarity between the inquired product and other recommended products may not impact the existence of the choice overload effect between conditions.

B.2. Time of Engagement

Note that while we can observe in our data when consumers received the messages and when they clicked on any product in the messages, we do not have the precise timestamps for when they *viewed* (i.e., opened) the messages. As a result, we are not able to directly analyze whether consumers' search probability and purchase probability vary with how quickly they view the message. However, for consumers who viewed the messages, if we assume that they decide immediately whether to click on any product after viewing the message, then the following equation holds ("min" stands for "minutes"):

$$\begin{aligned}
 & \text{Prob}(\text{click on any product within } X \text{ min}) \\
 = & \text{Prob}(\text{view message within } X \text{ min}) \times \text{Prob}(\text{click on any product} | \text{view message within } X \text{ min})
 \end{aligned}$$

Table A6 Comparing Choice Overload Effect Between Different Levels of Product Sub-category Alignment With Group 3 as the Baseline

Outcome Variable	SearchIncidence (1)	PurchaseIncidence (2)
Group 2	0.0166**** (0.0017)	0.0059**** (0.0009)
Group 4	-0.0034** (0.0016)	-0.0031**** (0.0009)
SameCategory	0.0091**** (0.0015)	0.0014* (0.0008)
Group 2 $\times$ SameCategory	-0.0016 (0.0019)	0.0002 (0.0011)
Group 4 $\times$ SameCategory	0.0013 (0.0019)	0.0012 (0.0010)
Merchant Fixed Effects	Yes	Yes
Date Fixed Effects	Yes	Yes
Observations	1,185,642	1,185,642

Notes: This table reports results from OLS regressions that predict search probability (Column (1)) and purchase probability (Column (2)) as a function of experimental groups (with Group 3 as the baseline), an indicator for whether the second recommended product is in the same sub-category as the inquired product, and the interactions. Standard errors are robust and clustered at the merchant level. The platform did not record the sub-categories of recommended products during the first three days of the experiment by accident. Thus, the number of observations in the regressions reported here is slightly smaller than that in our main analysis. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$; **** $p < 0.001$.

given any $X > 0$. In the above equation, Prob(view message within X min) should not be affected by the number of recommended products in the message, because consumers were only exposed to recommended products and were affected by the number of recommended products after they viewed a message. Therefore, though we cannot directly estimate how search probability varies with recommendation set size for consumers who engage with messages quickly (i.e., view them within just a couple of minutes), we can glean some insights by examining how the probability of consumers clicking on any product quickly (e.g., within just a couple of minutes) changes with the recommendation set size. The same idea applies to consumers who engage with the messages late (i.e., after a longer delay): We can examine how the probability of consumers clicking on any product after a few minutes changes with the recommendation set size. Following this idea, we compare all consumers who viewed a message across four experimental groups in terms of (1) their probability of clicking within 4.8 minutes and (2) the probability of clicking after 4.8 minutes (with 4.8 minutes being the median amount time it took consumers in our data to click on their recommendation message).

As shown in Figure A5, both the probabilities of consumers clicking on any product within 4.8 minutes and the probabilities of them clicking after 4.8 minutes exhibit an inverted-U shaped pattern as the number of recommended products increases. To provide statistical evidence for the inverted-U shape, we estimate the impacts of recommendation set size on the probability of consumers clicking more quickly or relatively more slowly. Using Specification (1) in the paper, we predict two binary outcome measures: (1) $SearchIncidenceQuick_{ijt}$, which equals 1 if a consumer clicked on the message within 4.8 minutes of receiving the recommendation message and 0 otherwise, and (2) $SearchIncidenceLate_{ijt}$, which equals 1 if a consumer clicked on the message after 4.8 minutes of receiving the recommendation message and 0 otherwise. The regression results, as shown in Table A7, support the pattern observed in Figure A5 whereby the proportions of consumers clicking on any product within 4.8 minutes and the proportions clicking after 4.8 minutes both

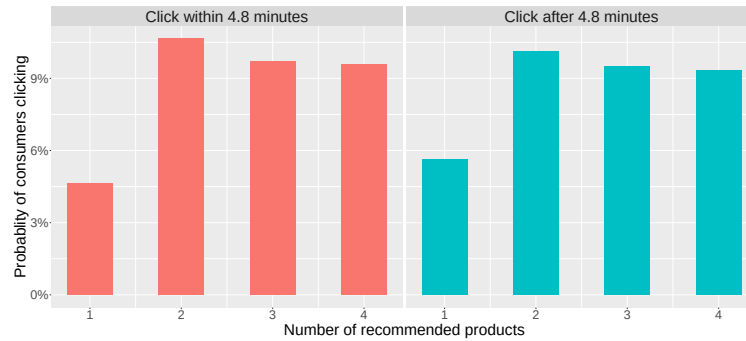


Figure A5 Probability of Consumers Clicking Within vs. After 4.8 Minutes

first increase as recommendation set size increases from 1 to 2 and then decrease as the recommendation set size increases from 2 to 3 or from 3 to 4. All the estimated treatment effects of increasing recommendation set size are statistically significant. While this is not a direct comparison of treatment effects at different engagement timing, the results provide suggestive evidence that the timing of consumers' engagement with the recommendation messages may not significantly influence our empirical results regarding the effect of recommendation set size on consumer search behavior.

Table A7 Treatment Effects on Consumer Clicking Quickly and Clicking Late

Outcome Variable	SearchIncidenceQuick		SearchIncidenceLate	
	Baseline: Group 2 (1)	Baseline: Group 3 (2)	Baseline: Group 2 (3)	Baseline: Group 3 (4)
Group 1	−0.0603*** (0.0006)	−0.0508*** (0.0006)	−0.0440*** (0.0006)	−0.0379*** (0.0006)
Group 2		0.0094*** (0.0006)		0.0061*** (0.0006)
Group 3	−0.0094*** (0.0006)		−0.0061*** (0.0006)	
Group 4	−0.0108*** (0.0006)	−0.0013** (0.0006)	−0.0075*** (0.0006)	−0.0014** (0.0006)
Merchant Fixed Effects	Yes	Yes	Yes	Yes
Date Fixed Effects	Yes	Yes	Yes	Yes
Observations	1,615,325	1,615,325	1,615,325	1,615,325

Notes: Standard errors are robust and clustered at the merchant level. Columns (1) and (2) report the treatment effects of recommending different number of products on probabilities of clicking quickly (i.e., within 4.8 minutes), with Group 2 and Group 3 as baseline groups, respectively. Columns (3) and (4) report similar effects on probabilities of clicking late (i.e., after 4.8 minutes) with the same moderator and baselines. * $p < 0.1$; ** $p < 0.05$; *** $p < 0.01$; **** $p < 0.001$.

Next, to alleviate the concerns regarding the selection bias in the analysis above, we can also focus on consumers who clicked on at least one product in the messages and compare the *purchase* behavior of consumers who clicked sooner to those who clicked later. We divide the consumers who clicked on at least one product into two sub-samples: (1) those who clicked within 4.8 minutes of receiving a message, and (2) those who clicked after 4.8 minutes. Figure A6 shows that in these two consumer sub-samples, purchase probabilities (conditional on clicking at least one product) similarly decrease as the number of recommended products increases. This correlational observation provides suggestive evidence that the impact of the number of recommended products on consumer purchase behavior may not vary significantly with the consumers' response time.

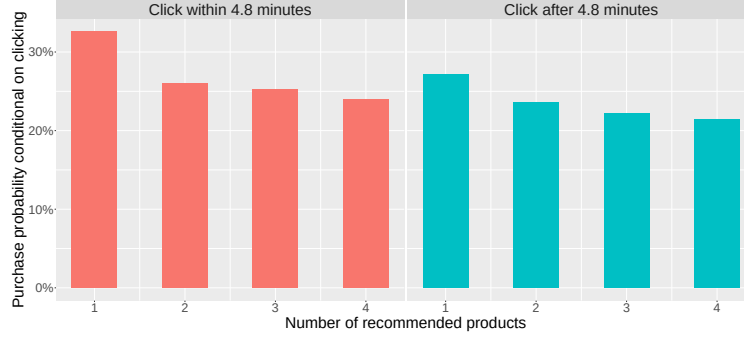


Figure A6 Purchase Probability Conditional on Clicking Among Consumers Who Clicked Within vs. After 4.8 Minutes

B.3. Comparison of Consumers in the Experiment with General Consumer Population

We obtained additional data from Alibaba Group and compared the pre-experiment purchasing behavior of consumers in our experiment against that of the general consumer population. Specifically, we randomly sampled 1 million consumers who made a purchase within one month before the experiment (hereafter referred to as “pre-experiment period”). Then for the consumers in our experiment and the 1 million sampled general consumers (which we hereafter collectively refer to as “consumers in our comparison analysis”), we collected their order data during the pre-experiment period. Figures A7(a) and A7(b) present the distributions of number of orders and expenditure in this period for those two samples of consumers.²⁴ The figures demonstrate that the distributions have similar modes and are similarly skewed. Specifically, in both consumer populations, the modal number of orders (after scaling) is around 0.05, and the modal order amount (after scaling) is around 0.01.²⁵ Moreover, the distributions are right-skewed (i.e., the mean is greater than the median) for both consumer populations. We do notice that a larger proportion of consumers in our experiment have high order frequencies and/or high expenditures, compared to the general consumer population. Specifically, 62.7% of consumers in the experiment have a purchase frequency higher than or equal to the median of the general consumers, and 60.5% have an order amount higher than or equal to the median of the general consumers. Put together, the comparison suggests that while we may have moderately oversampled the more active purchasers on the platform in our experiment, the consumers in our experiment largely represent the general consumer population in terms of diversity in purchasing behavior. These findings, combined with the fact that a significant proportion (35%) of consumers who purchase on the platform make an inquiry before their purchase, indicate that our empirical findings can be quite relevant for the platform.

²⁴ For data confidentiality reasons, we have to scale the data rather than reporting the raw values. Specifically, we scale the minimum values of both variables (the number of orders and average order amount) to 0. Given that the original distributions of both variables are extremely right-skewed, we scale the 95% percentiles of the original distributions to 1 for both variables and only display the distributions of the scaled data up to 1 in Figure A7 for better visual clarity and interpretability.

²⁵ The statistics are computed using the full sample of consumers in our comparison analysis with values scaled in the same way as described in Footnote 23.

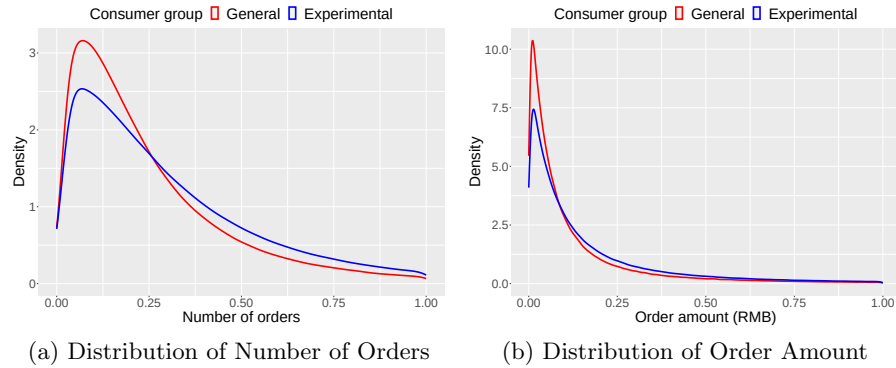


Figure A7 Comparison of Consumers in Experiment and General Consumers in Purchasing Behavior During the Month Prior to the Experiment

C. Theoretical Model

In this appendix, we present a simple two-stage model to illustrate the predictions of a classic choice model versus those with anticipated regret. We assume the same marginal search cost for every product for simplicity, but our model results are robust if we assume a very low search cost or very low value uncertainty for the first product (which may be the case if it is a product consumers have interacted with before). Our model can also be extended to capture additional behavioral effects of the recommendation message that are invariant in set size. For instance, recommendation messages, if used as part of a retargeting strategy, may serve as a shopping reminder for consumers. Since this effect would arguably be fixed for all recommendation set sizes, we can include this effect as a reduction in the fixed cost of starting a search (i.e., ε_{NS} in the model that we introduce below). In addition, our model can be extended to incorporate an additional fixed utility of search for $n \geq 2$ to represent the difference between $n = 1$ (recommending one product that customers have previously interacted with) and $n \geq 2$ (additionally recommending new products). These changes would not qualitatively change our main model insights (i.e., that search and purchase probabilities increase in n in the normative setting, but may decrease in n with the presence of anticipated regret).

C.1. Two-stage Setting

Suppose a consumer is presented with n substitutable products, denoted by $i = 1, \dots, n$. We refer to $S = \{1, 2, \dots, n\}$ as the choice set. Ex ante, each option's value v_i consists of a known part μ_i and an uncertain part ϵ_i , i.e., $v_i = \mu_i + \epsilon_i$. The uncertainty reflects consumers' imperfect information about product attributes and about the fit between the product and the consumer's idiosyncratic preferences. We assume that ϵ_i 's are identically and independently distributed, and that $E[\epsilon_i] = 0$ for all $i \in S$. Denote the no-purchase option by 0, and let its value be given by v_0 , which is known to the consumer. For ease of exposition and without loss of generality, we normalize price to zero. The value v_i can also be interpreted as the consumer's surplus (net product price) from purchasing product i .

We suppose that consumers follow a two-stage decision process to make their purchase decisions, which is similar to the *consider-then-choose* choice models that have been studied in the literature (Hauser 2014,

Wang and Sahin 2018).²⁶ In the first stage, the consumer decides whether to search any product and, if so, which product(s) to search. Let S_D represent the set of products that the consumer chooses to search, namely the consideration set, with $S_D \subseteq S$. In the second stage, the consumer learns the value v_i of every option in her consideration set $i \in S_D$. The consumer then chooses an option from $S_D^+ := S_D \cup \{0\}$. The consumer cannot purchase a product without searching it first. Consistent with prior literature, we assume that the consumer expends a positive search cost c for each product that she searches to learn its value.

A rational consumer's decision-making problem can be characterized as follows. In the second stage, given a consideration set S_D , the consumer chooses the option i^* in S_D^+ that yields the highest value, i.e.,

$$i^* = \arg \max_{i \in S_D^+} v_i.$$

Correspondingly, in the first stage, the consumer's expected utility from searching a consideration set S_D (*prior* to learning the values of the options in S_D) is

$$\mathbb{E}U(S_D) = \mathbb{E}[\max_{i \in S_D^+} v_i] - c|S_D|. \quad (\text{A1})$$

If the consumer decides to search in the first stage, she would choose the consideration set S_D^* that maximizes her expected utility given in (A1), i.e.,

$$S_D^* = \arg \max_{S_D \subseteq S, S_D \neq \emptyset} \mathbb{E}U(S_D).$$

The consumer may also decide not to search any product, which yields a utility of $U(\emptyset) = v_0$.

To model variability in the search decision, we apply a random noise to the consumer's search decision, so that the consumer searches if and only if $\mathbb{E}U(S_D^*) \geq v_0 + \varepsilon_{NS}$. The random term ε_{NS} can be interpreted as idiosyncratic shocks to the consumer's outside option of no-search or to the fixed cost of starting a search (Pires 2016). Therefore, the consumer's probability of searching at least one product from the choice set S is given by

$$\psi(S) = \Pr(\varepsilon_{NS} \leq \mathbb{E}U(S_D^*) - v_0). \quad (\text{A2})$$

The probability of a consumer purchasing at least one product from the choice set S unconditional on search, which we denote as $\omega(S)$, is the product of the consumer's search probability $\psi(S)$ and her purchase probability conditional on search. That is,

$$\omega(S) = \psi(S) \cdot \Pr(\max_{i \in S_D^+} v_i \geq v_0).$$

Finally, consistent with traditional recommender systems as well as our field experiment setting, we assume that products are ranked by preference weights when added into the choice set:

ASSUMPTION 1. *For any two choice sets S and S' where $|S| < |S'|$, we have $S \subset S'$ and $\max\{\mu_i : i \in S' \setminus S\} \leq \min\{\mu_j : j \in S\}$.*

²⁶ We assume that consumers engage in simultaneous search in our model. Several empirical studies have analyzed consumer shopping data and found that the observed shopping behavior is more consistent with simultaneous—as opposed to sequential—search strategies (De los Santos et al. 2012, Honka and Chintagunta 2017).

Assumption 1 states that for any two choice sets with different sizes, the smaller one is always a subset of the larger one. This matches our experiment setting and is also the typical setup in existing studies on the choice overload effect.²⁷ Assumption 1 also states that the expected values of the additional products in the larger set do not exceed those of the products in the smaller set. In other words, as the size of the choice set increases, the firm adds new products with expected values smaller than those of the existing ones. This is a common feature of recommender systems like the one employed in our field experiment, as we explain in more detail in Section 4.

Proposition 1 describes the rational consumer's behavior as choice set size increases. Throughout this appendix, all monotonicity descriptions refer to weak monotonicity, unless stated otherwise.

PROPOSITION 1. *As the size of the choice set S increases, the rational consumer's search probability $\psi(S)$ and purchase probability $\omega(S)$ both increase.*

See Appendix D for all proofs. Proposition 1 states that as choice set size increases, a rational consumer should always be more likely to start a search or to make a purchase. This result reflects the intuition that at each stage, the probability of choosing the outside option (i.e., no search or no purchase) should never increase as the set of available options expands.

C.2. Anticipated Regret

Prior literature has not addressed the implications of the anticipated regret mechanism for consumers' search and purchase behavior separately. Therefore, to generate predictions, we formally model the mechanism in our two-stage setting below.²⁸ Following Bell (1982) and Loomes and Sugden (1982), we assume that if the consumer purchases an ex post sub-optimal option, it generates a regret disutility that is proportional to the difference between the best alternative option's value and the consumer's actual choice value. Anticipating this, the consumer's utility from choosing option $i \in S_D$ in the second stage is given by

$$\tilde{v}_i(S_D) = v_i - \beta \mathbb{E} \left[(\max_{j \in S^+} v_j - v_i)^+ \mid \{v_k : k \in S_D^+\} \right], \quad i \in S_D, \quad (\text{A3})$$

where $S^+ := S \cup \{0\}$. The second term in Equation (A3) represents the consumer's expected disutility from regret. The parameter $\beta \geq 0$ captures the extent to which anticipated regret influences the consumer's decisions, and the term $\mathbb{E} [(\max_{j \in S^+} v_j - v_i)^+ \mid \{v_k : k \in S_D^+\}]$ can be interpreted as the consumer's expected foregone surplus, where $\{v_k : k \in S_D^+\}$ is the information set available in the second stage (i.e., the observed values of the options searched) and the expectation is taken with respect to the values of the options NOT searched, i.e., $\{v_j : j \in S \setminus S_D\}$. We assume that if the consumer does not make a purchase, then she does not anticipate any regret, i.e., $\tilde{v}_0(S_D) = v_0$. This is consistent with psychology research showing that people tend

²⁷ One can easily construct examples of the choice overload effect when this assumption does not hold. However, if the small choice set is not a subset of the large one, then the regularity property no longer applies and there could be rational explanations for why consumers may be more likely to search and purchase from the small choice set than the large one.

²⁸ Our work extends the model in Irons and Hepburn (2007), which (1) assumes that consumers who search always purchase and (2) only analyzes a special case of uniformly distributed values. Sarver (2008) also models consumer regret when choosing from a choice set, but does not allow for consumer search (so uncertainties cannot be resolved before purchase).

to feel more regret towards actions as opposed to inactions in purchase settings (Simonson 1992, Abendroth and Diehl 2006).²⁹

The consumer chooses the option i^* in S_D^+ that maximizes her expected utility, i.e., $i^* = \arg \max_{i \in S_D^+} \tilde{v}_i(S_D)$. Note that if the consumer searched all products in S , then she would choose the option with the highest value for certain, i.e., $v_{i^*} = \max_{j \in S^+} v_j$, and there is no anticipated regret associated with her optimal choice, i.e., $\tilde{v}_{i^*}(S) = v_{i^*}$. By contrast, if the consumer only searched a subset of S (i.e., $S_D \subsetneq S$) and finds it optimal to make a purchase, we can show that $i^* = \arg \max_{i \in S_D} v_i$, and the utility associated with her optimal choice can be expressed as

$$\tilde{v}_{i^*}(S_D) = v_{i^*} - \beta \mathbb{E} \left[\left(\max_{j \in S \setminus S_D} v_j - v_{i^*} \right)^+ \middle| v_{i^*} \right], \quad (\text{A4})$$

where $i^* \neq 0$.³⁰ That is, the consumer incurs anticipated regret which depends on the expected positive difference between the best alternative in the set of unsearched products (i.e., $S \setminus S_D$) and the consumer's optimal choice i^* from S_D .

In the first stage, the consumer decides whether to search any product and if so, which product(s) to search, by comparing the expected utility from each action. Specifically, her expected utility from searching a nonempty consideration set S_D is

$$\mathbb{E}U(S_D) = \mathbb{E} \left[\max_{i \in S_D^+} \tilde{v}_i(S_D) \right] - c|S_D|, \quad (\text{A5})$$

where $U(S_D) := \max_{i \in S_D^+} \tilde{v}_i(S_D) - c|S_D|$ is the utility of searching the consideration set S_D . As in the rational model, we denote the consumer's optimal consideration set given that she decides to search as $S_D^* = \arg \max_{S_D \subseteq S, S_D \neq \emptyset} \mathbb{E}U(S_D)$, and let the consumer's probability of search be represented by $\psi(S) = \Pr(v_0 + \epsilon_{NS} \leq \mathbb{E}U(S_D^*))$. The consumer's probability of making a purchase is then given by $\omega(S) = \psi(S) \cdot \Pr(\max_{i \in S_D^*} \tilde{v}_i(S_D^*) \geq v_0)$.

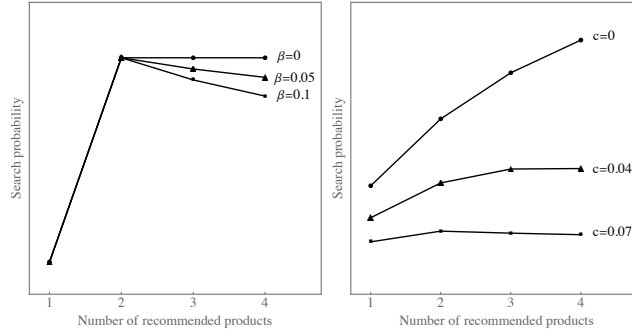
We can now derive the following predictions on consumers' search behavior.

PROPOSITION 2. (Search Probability) *Suppose $c > 0$ and $\beta > 0$. There exists a positive integer $\bar{n}$ such that as the size of the choice set S increases, the consumer's probability of searching at least one product, i.e., $\psi(S)$, first increases (when $|S| \leq \bar{n}$) then decreases (when $|S| \geq \bar{n}$).*

Note that if $\beta = 0$, then this model reverts back to the rational case, where we have shown that consumers always search (weakly) more when choice set size is larger. When there is regret (i.e., $\beta > 0$), by contrast, Proposition 2 shows that the consumer's probability of starting a search could first increase then decrease as choice set size increases. This is because the consumer incurs a disutility for not searching all available options; in this case, a larger choice set could reduce the consumer's expected utility of search due to either larger anticipated regret (if she does not search more options) or larger search costs (if she does search more options). As a result, the consumer may become less inclined to start a search in the first place. In fact, this effect is stronger when the consumer's anticipated regret parameter (β) is larger, as indicated in the following sensitivity result.

²⁹ Our theoretical results are not necessarily restricted by these assumptions. Similar results would hold under the alternative assumption that consumers do anticipate regret even if they do not purchase.

³⁰ See proof of Lemma 2 for more details.



Note: In this numerical simulation, we let $v_i \sim \text{Uniform}[0, 1]$ for $i \in S$, $v_0 = 0.5$, $\epsilon_{NS} \sim \text{Uniform}[0, 1]$, $c = 0.06$ in the left panel, and $\beta = 0.1$ in the right panel.

Figure A8 Search Probability As a Function of Recommendation-Set Size

COROLLARY 1. *As the choice set size increases from $\bar{n}$ to $\bar{n} + 1$, the reduction in a consumer's search probability $\psi(S)$ is larger when β is larger.*

Similarly, we can show numerically that for two choice sets S and S' where $|S| = \bar{n}$ and $|S'| = \bar{n} + 1$, as search cost c decreases, the difference in search probability $\psi(S) - \psi(S')$ decreases. Figure A8 illustrates a consumer's search probability as a function of the choice set size for different values of regret parameter β and search cost c .

Next we analyze the effect of choice set size on consumers' purchase behavior. Recall that the consumer's purchase probability is the product of her search probability and her likelihood of making a purchase given that she searches. Therefore, as choice set size increases, a reduction in search probability could drive a reduction in purchase probability, provided that the purchase probability conditional on search also decreases (or does not increase too much). The next proposition shows that this is the case when the regret parameter is not too large.

PROPOSITION 3. (Purchase Probability) *Suppose $c > 0$ and $\beta > 0$. There exists some $\bar{\beta}$ such that when $\beta < \bar{\beta}$, as the size of the choice set S increases, the consumer's probability of purchasing at least one product, i.e., $\omega(S)$, first increases (when $|S| \leq \bar{n}$) then decreases (when $|S| \geq \bar{n}$), where $\bar{n}$ is as given in Proposition 2.*

Proposition 3 thus provides sufficient conditions for the choice overload effect to exist.³¹ Overall, our analysis suggests that in the presence of search cost and with moderate parameter values for anticipated regret, both search and purchase probabilities could first increase then decrease as choice set size increases.

D. Proofs

Throughout the proofs, we assume without loss of generality that the products in any choice set S are indexed such that the expected values of the products are nonincreasing in the index, i.e., $\mu_1 \geq \mu_2 \geq \dots \geq \mu_n$ for any $S = \{1, 2, \dots, n\}$. Moreover, we denote $A^+ := A \cup \{0\}$ for any set A . Define $\mathbf{1}_A$ as an indicator function which equals 1 if condition A holds and 0 otherwise.

³¹ See the proof of Proposition 3 for detailed characterization of the threshold $\bar{\beta}$.

Proof of Proposition 1. Consider two choice sets S and S' where $|S| < |S'|$. By Assumption 1, the small choice set is a subset of the large one, i.e., $S \subset S'$. We will prove $\psi(S) \leq \psi(S')$ and $\omega(S) \leq \omega(S')$. Denote the optimal non-empty consideration set under S and S' by $S_D^*(S)$ and $S_D^*(S')$, respectively.

To prove $\psi(S) \leq \psi(S')$, it is sufficient to prove $\mathbb{E}U(S_D^*(S)) \leq \mathbb{E}U(S_D^*(S'))$, where $\mathbb{E}U(S_D^*(S)) = \max_{S_D \subseteq S, S_D \neq \emptyset} \mathbb{E}U(S_D)$. That is, the maximum expected utility from starting a search is weakly lower under S than under S' . This inequality holds because $S_D^*(S) \subseteq S$ and $S \subset S'$. In other words, the maximum expected utility under S (achieved by searching the subset $S_D^*(S)$) is always achievable under S' .

Next, to prove $\omega(S) \leq \omega(S')$, it suffices to prove $\Pr(\max_{i \in S_D^*(S)} v_i \geq v_0) \leq \Pr(\max_{i \in S_D^*(S')} v_i \geq v_0)$, which holds if the optimal consideration set under S is a subset of the optimal consideration set under S' (i.e., $S_D^*(S) \subseteq S_D^*(S')$).

Suppose the sizes of $S_D^*(S)$ and $S_D^*(S')$ are t and t' , respectively. Under the i.i.d. assumption for ε_i 's, $S_D^*(S)$ must contain the t products with the highest expected values in the choice set S , i.e., $S_D^*(S) = \{1, 2, \dots, t\}$. Similarly, $S_D^*(S')$ must contain the t' products with the highest expected values in S' . By Assumption 1, $\{1, 2, \dots, t\}$ are the t products with the highest expected values in both choice sets S and S' . Therefore, we have either (I) $S_D^*(S) \subseteq S_D^*(S')$ (if $t \leq t'$) or (II) $S_D^*(S') \subset S_D^*(S)$ (if $t' < t$), and it suffices to prove that case (II) does not exist. Note that since $S \subset S'$, both $S_D^*(S)$ and $S_D^*(S')$ are subsets of S' , and the optimality condition of $S_D^*(S')$ requires that $\mathbb{E}U(S_D^*(S')) \geq \mathbb{E}U(S_D^*(S))$. The previous inequality can only hold if $S_D^*(S) \subseteq S_D^*(S')$ (because otherwise $S_D^*(S') \subset S_D^*(S)$ and hence, by the optimality condition of $S_D^*(S)$, $\mathbb{E}U(S_D^*(S')) < \mathbb{E}U(S_D^*(S))$). $\square$

We establish the following two lemmas before presenting the proofs for Propositions 2 and 3.

LEMMA 1. Let $S_n = \{1, 2, \dots, n\}$ denote the choice set of size n , and define $V_n = \max_{i \in S_n^+} v_i$. Then $\mathbb{E}[V_n]$ is increasing concave in n . Furthermore, $\mathbb{E}[V_n] - \mathbb{E}[V_{n-1}] \rightarrow 0$ as $n \rightarrow \infty$.

Proof of Lemma 1. See the Electronic Companion. $\square$

LEMMA 2. Given the regret parameter $\beta > 0$ and the choice set S , the optimal consideration set is $S_D^* = \{1, 2, \dots, d\}$ for some integer d where $1 \leq d \leq |S|$.

Proof of Lemma 2. See the Electronic Companion. $\square$

Proof of Proposition 2. Let $S_D^*(S_n)$ denote the optimal nonempty consideration set when a consumer is offered the choice set $S_n = \{1, 2, \dots, n\}$. Define $U_n := U(S_D^*(S_n))$ as the consumer's maximum expected utility from searching at least one product in S_n . By Assumption 1 and Lemma 2, the optimal consideration set $S_D^*(S_n)$ must be $S_m = \{1, 2, \dots, m\}$ for some integer m where $1 \leq m \leq n$. Let $U_{n,m}$ be the consumer's expected utility of searching the m products with the highest expected values (i.e., S_m) from the choice set S_n . By proof of Lemma 2, we have $U_{n,m} = \mathbb{E} \max_{i \in S_m^+} \hat{v}_i(n, m) - cm$, where $\hat{v}_i(n, m)$ is defined as

$$\hat{v}_i(n, m) = \begin{cases} v_i - \mathbf{1}_{i \neq 0} \cdot \beta \mathbb{E} \left[\left(\max_{j \in S_n \setminus S_m} v_j - v_i \right)^+ \middle| v_i \right], & i \in S_m^+ \text{ where } m < n, \\ v_i, & i \in S_m^+ \text{ where } m = n. \end{cases}$$

We first prove the following useful result.

Result 1: For a choice set $S_n = \{1, 2, \dots, n\}$ with $n > 1$, if the consumer prefers partial search over full search (i.e., $|S_D^*(S_n)| < n$), then we have $\mathbb{E}U_n \leq \mathbb{E}U_{n-1}$.

Proof of Result 1: If the consumer prefers partial search over full search for S_n , this implies $\mathbb{E}U_n = \max\{U_{n,1}, \dots, U_{n,n-1}\}$. We have

$$\mathbb{E}U_n = \max\{U_{n,1}, \dots, U_{n,n-1}\} \leq \max\{U_{n-1,1}, \dots, U_{n-1,n-1}\} = \mathbb{E}U_{n-1}.$$

The above inequality holds because for each $m \in \{1, \dots, n-1\}$, we have

$$\begin{aligned} U_{n,m} &= \mathbb{E} \max_{i \in S_m^+} (v_i - \mathbf{1}_{i \neq 0} \cdot \beta \mathbb{E}[(\max_{j \in S_n \setminus S_m} v_j - v_i)^+ | v_i]) - cm \\ &\leq \mathbb{E} \max_{i \in S_m^+} (v_i - \mathbf{1}_{i \neq 0} \cdot \beta \mathbb{E}[(\max_{j \in S_{n-1} \setminus S_m} v_j - v_i)^+ | v_i]) - cm \\ &= U_{n-1,m}, \end{aligned}$$

where the inequality holds because $\beta > 0$ and given any value of v_i where $i \in S_m$, $\mathbb{E}[(\max_{j \in S_{n-1} \setminus S_m} v_j - v_i)^+] \leq \mathbb{E}[(\max_{j \in S_n \setminus S_m} v_j - v_i)^+]$ (since $S_{n-1} \setminus S_m \subset S_n \setminus S_m$). ■

The main proof is by forward induction. Suppose that the consumer prefers full search over partial search for S_n , i.e., $\mathbb{E}U_n = \max\{U_{n,1}, \dots, U_{n,n}\} = U_{n,n}$. We can prove that if $\mathbb{E}U_{n+1} \leq \mathbb{E}U_n$, then $\mathbb{E}U_{N+1} \leq \mathbb{E}U_N$ for all $N \geq n+1$. That is, if adding one product to the choice set decreases the consumer's expected utility of search, then adding more products will continue to decrease the expected utility of search.

To show this, we observe that the conditions $\mathbb{E}U_n = U_{n,n}$ and $\mathbb{E}U_{n+1} \leq \mathbb{E}U_n$ imply $U_{n+1,n+1} \leq \mathbb{E}U_{n+1} \leq \mathbb{E}U_n = U_{n,n}$. As a result of the concavity of $U_{n,n}$ in n (by Lemma 1 and noting that $U_{n,n} = \mathbb{E}V_n - cn$), it follows that $U_{N+1,N+1} \leq U_{N,N}$ for all $N \geq n+1$. Thus for $N \geq n+1$, we have

$$U_{N+1,N+1} \leq U_{N,N} \leq \max\{U_{N,1}, \dots, U_{N,N}\} = \mathbb{E}U_N.$$

As in the proof of Result 1, we also have

$$\max\{U_{N+1,1}, \dots, U_{N+1,N}\} \leq \max\{U_{N,1}, \dots, U_{N,N}\} = \mathbb{E}U_N.$$

Combining the two inequalities above yields

$$\mathbb{E}U_{N+1} = \max\{U_{N+1,1}, \dots, U_{N+1,N+1}\} \leq \max\{U_{N,1}, \dots, U_{N,N}\} = \mathbb{E}U_N.$$

Finally, define $\bar{n} = \arg \max_n U_{n,n}$ (where we select the largest value to break ties). By concavity of $U_{n,n}$ (see discussion above), $U_{n,n}$ is increasing in n when $n \leq \bar{n}$. Thus by Result 1, $\mathbb{E}U_n = U_{n,n}$ and $\mathbb{E}U_n$ is increasing in n when $n \leq \bar{n}$ (this is because if there exists $n \leq \bar{n}$ such that $\mathbb{E}U_n > U_{n,n}$, then Result 1 implies that $\mathbb{E}U_n \leq \mathbb{E}U_{n-1}$, which implies $\mathbb{E}U_{n-1} > U_{n,n} \geq U_{n-1,n-1}$, and by induction, $\mathbb{E}U_1 > U_{1,1}$; a contradiction). Next we show $\mathbb{E}U_{\bar{n}} \geq \mathbb{E}U_{\bar{n}+1}$. This holds because if $\mathbb{E}U_{\bar{n}+1} = U_{\bar{n}+1,\bar{n}+1}$, then by definition of $\bar{n}$, $\mathbb{E}U_{\bar{n}} = U_{\bar{n},\bar{n}} > U_{\bar{n}+1,\bar{n}+1} = \mathbb{E}U_{\bar{n}+1}$. If $\mathbb{E}U_{\bar{n}+1} > U_{\bar{n}+1,\bar{n}+1}$, then $|S_D^*(S_{\bar{n}+1})| < \bar{n} + 1$ and again we have $\mathbb{E}U_{\bar{n}} \geq \mathbb{E}U_{\bar{n}+1}$ by Result 1. Thus by our proof above, $\mathbb{E}U_n$ is decreasing in n when $n \geq \bar{n}$. We can conclude that there exists a threshold $\bar{n}$ such that $\mathbb{E}U_n$ increases in n for $n \leq \bar{n}$ then decreases in n otherwise. Note that $\bar{n}$ is finite because $\mathbb{E}U_{n+1} \leq \mathbb{E}U_n$ for $n \rightarrow +\infty$. To see this, observe that if $\mathbb{E}U_{n+1} > \mathbb{E}U_n$ for $n \rightarrow +\infty$, then we must have $\mathbb{E}U_n = U_{n,n}$ and $\mathbb{E}U_{n+1} = U_{n+1,n+1}$ (Result 1), which means $U_{n+1,n+1} > U_{n,n}$ for $n \rightarrow +\infty$. However, by Lemma 1 we know $U_{n+1,n+1} < U_{n,n}$ for $n \rightarrow +\infty$; a contradiction.

This concludes the proof. □

Proof of Corollary 1. By the proof of Proposition 2, when the choice set size is $\bar{n}$, the consumer finds it optimal to conduct full search, in which case the expected utility from starting a search is $\mathbb{E}U_{\bar{n}} = U_{\bar{n},\bar{n}} = \mathbb{E}[V_{\bar{n}}] - c\bar{n}$. Since $\mathbb{E}U_{\bar{n}}$ is independent of β and $\mathbb{E}U_{\bar{n}+1}$ is (weakly) decreasing in β , it follows that the difference in search probability, i.e., $\psi(S_{\bar{n}}) - \psi(S_{\bar{n}+1})$, is larger when β increases. $\square$

Proof of Proposition 3. See the Electronic Companion. $\square$